A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer's disease

Douglas P. Wightman[1], Iris E. Jansen[1], Jeanne E. Savage[1], Alexey A. Shadrin[2,3], Shahram Bahrami[2–4], Dominic Holland[5], Arvid Rongve[6,7], Sigrid Børte[3,8,9], Bendik S. Winsvold[9–11], Ole Kristian Drange[12,13], Amy E. Martinsen[3,9,10], Anne Heidi Skogholt[9,14], Cristen Willer[15], Geir Bråthen[16–18], Ingunn Bosnes[12,19], Jonas Bille Nielsen[9,15,20], Lars G. Fritsche[21], Laurent F. Thomas[9,14], Linda M. Pedersen[10], Maiken E. Gabrielsen[9], Marianne Bakke Johnsen[3,8,9], Tore Wergeland Meisingset[16,17], Wei Zhou[22,23], Petroula Proitsi[24], Angela Hodges[24], Richard Dobson[25–29], Latha Velayudhan[24], 23andMe Research Team[30], Julia M. Sealock[31,32], Lea K. Davis[31,32], Nancy L. Pedersen[33], Chandra A. Reynolds[34], Ida K. Karlsson[33,35], Sigurdur Magnusson[36], Hreinn Stefansson[36], Steinunn Thordardottir[37], Palmi V. Jonsson[37,38], Jon Snaedal[37], Anna Zettergren[39], Ingmar Skoog[39,40], Silke Kern[39,40], Margda Waern[39,41], Henrik Zetterberg[42–45], Kaj Blennow[44,45], Eystein Stordal[12,19], Kristian Hveem[9,46], John-Anker Zwart[3,9,10], Lavinia Athanasiu[2,4], Per Selnes[47], Ingvild Saltvedt[16,18], Sigrid B. Sando[16,17], Ingun Ulstein[48], Srdjan Djurovic[49,50], Tormod Fladby[3,47], Dag Aarsland[24,51], Geir Selbæk[3,48,52], Stephan Ripke[23,53,54], Kari Stefansson[36], Ole A. Andreassen[2–4*], Danielle Posthuma[1,55*†]

1. *Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Amsterdam Neuroscience, VU University Amsterdam, The Netherlands*.
2. *NORMENT Centre, University of Oslo, Oslo, Norway*.
3. *Institute of Clinical Medicine, University of Oslo, Oslo, Norway.*
4. *Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway.*
5. *Department of Neurosciences, University of California, San Diego, La Jolla, CA 92037, USA*.
6. *Department of Research and Innovation, Helse Fonna, Haugesund Hospital, Haugesund, Norway.*
7. *The University of Bergen, Institute of Clinical Medicine (K1), Bergen Norway*.
8. *Research and Communication Unit for Musculoskeletal Health (FORMI), Department of Research, Innovation and Education, Division of Clinical Neuroscience, Oslo University Hospital, Oslo, Norway.*
9. *K. G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Trondheim, Norway.*
10. *Department of Research, Innovation and Education, Division of Clinical Neuroscience, Oslo University Hospital, Oslo, Norway.*
11. *Department of Neurology, Oslo University Hospital, Oslo, Norway.*
12. *Department of Mental Health, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Trondheim, Norway.*
13. *Division of Mental Health Care, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway.*
14. *Department of Clinical and Molecular Medicine, Norwegian University of Science and Technology, Trondheim, Norway.*
15. *Department of Internal Medicine, Division of Cardiovascular Medicine, University of Michigan, Ann Arbor, 48109, MI, USA*.
16. *Department of Neuromedicine and Movement Science, Norwegian University of Science and Technology, Trondheim, Norway*.
17. *Department of Neurology and Clinical Neurophysiology, University Hospital of Trondheim, Norway.*
18. *Department of Geriatrics, St. Olav's Hospital, Trondheim University Hospital, Norway*.
19. *Department of Psychiatry, Hospital Namsos, Nord-Trøndelag Health Trust, Namsos, Norway.*
20. *Department of Epidemiology Research, Statens Serum Institut, Copenhagen, Denmark.*
21. *Center for Statistical Genetics, Department of Biostatistics, University of Michigan, Ann Arbor, 48109, MI, USA.*
22. *Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA*.

23. *Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts, USA.*
24. *Institute of Psychiatry, Psychology and Neurosciences, King's College London*.
25. *Department of Biostatistics and Health Informatics, Institute of Psychiatry, Psychology and Neuroscience (IoPPN), King's College London, 16 De Crespigny Park, London, SE5 8AF, UK*.
26. *NIHR Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London, London, UK*.
27. *Health Data Research UK London, University College London, London, UK*.
28. *Institute of Health Informatics, University College London, London, UK*.
29. *NIHR Biomedical Research Centre at University College London Hospitals NHS Foundation Trust, London, UK*.
30. *23andMe, Inc., Mountain View, CA, USA*.
31. *Division of Genetic Medicine, Department of Medicine Vanderbilt University Medical Center Nashville, TN, 37232, USA*.
32. *Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN, 37232, USA*.
33. *Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden*.
34. *Department of Psychology, University of California-Riverside, Riverside, CA, USA*.
35. *Institute of Gerontology and Aging Research Network – Jönköping (ARN-J), School of Health and Welfare, Jönköping University, Jönköping, Sweden*.
36. *deCODE Genetics/Amgen, Sturlugata 8, IS-101, Reykjavik, Iceland*.
37. *Department of Geriatric Medicine, Landspitali University Hospital, Reykjavik, Iceland*.
38. *Faculty of Medicine, University of Iceland*.
39. *Neuropsychiatric Epidemiology Unit, Department of Psychiatry and Neurochemistry, Institute of Neuroscience and Physiology, the Sahlgrenska Academy, Centre for Ageing and Health (AGECAP) at the University of Gothenburg, Sweden.*
40. *Region Västra Götaland, Sahlgrenska University Hospital, Psychiatry, Cognition and Old Age Psychiatry Clinic, Gothenburg, Sweden*.
41. *Region Västra Götaland, Sahlgrenska University Hospital, Psychosis Clinic, Gothenburg, Sweden.*
42. *Department of Neurodegenerative Disease, UCL Institute of Neurology, London, United Kingdom.*
43. *UK Dementia Research Institute at UCL, London, United Kingdom.*
44. *Department of Psychiatry and Neurochemistry, Institute of Neuroscience and Physiology, the Sahlgrenska Academy at the University of Gothenburg, Mölndal, Sweden.*
45. *Clinical Neurochemistry Laboratory, Sahlgrenska University Hospital, Mölndal, Sweden.*
46. *HUNT Research Center, Department of Public Health and Nursing, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Trondheim, Norway.*
47. *Department of Neurology, Akershus University Hospital, Lørenskog, Norway*.
48. *Department of Geriatric Medicine, Oslo University Hospital, Oslo, Norway*.
49. *Department of Medical Genetics, Oslo University Hospital, Oslo, Norway*.
50. NORMENT, Department of Clinical Science, University of Bergen, Bergen, Norway.
51. *Centre of Age-Related Medicine, Stavanger University Hospital, Norway*.
52. *Norwegian National Advisory Unit on Ageing and Health, Vestfold Hospital Trust, Tønsberg, Norway.*
53. Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA.
54. *Department of Psychiatry and Psychotherapy, Charité–Universitätsmedizin, Berlin, Germany*.
55. *Department of Child and Adolescent Psychiatry and Pediatric Psychology, Section*

Complex Trait Genetics, Amsterdam Neuroscience, Vrije Universiteit Medical Center, Amsterdam University Medical Center, Amsterdam, The Netherlands.
*These authors contributed equally to this work
†Correspondence should be addressed to: Danielle Posthuma: Department of Complex Trait Genetics, VU University, De Boelelaan 1085, 1081 HV, Amsterdam, The Netherlands. Phone: +31 20 598 2823, Fax: +31 20 5986926, d.posthuma@vu.nl

Table of Contents

Supplementary

Supplementary Results

Comparison to Jansen *et al.* (2019)

In comparison to our previous meta-analysis[1], we failed to replicate the association of *ADAMTS4*, *HESX1*, *CNTNAP2*, *KAT8*, *SCIMP*, *ALPK2*, and *AC074212.3* in the current study (**Supplementary Table 5**). The lead variants in *ADAMTS4*, *KAT8*, *SCIMP*, and *ALPK2* all had P-values smaller than $3x10^{-6}$ (**Supplementary Figures 11-14**), there was relatively small support for the lead variant in *AC074212.3* ($P=2.96x10^{-5}$) with the majority of the support coming from the UK Biobank (UKB) ($P=1.80x10^{-5}$) and IGAP ($P=1.22x10^{-4}$) datasets (**Supplementary Figure 15**). The lead variants for *HESX1* and *CNTNAP2* identified in Jansen *et al.* (2019)[1] were rare variants so they were largely excluded from the meta-analysis (except the lead variant of *HESX1* which was present in the 23andMe dataset; $P=0.70$). The failure to replicate these loci can be attributed to the low significance of the lead variants in the datasets which are included in our study but not included in Jansen et al. (2019)[1] (Finngen, GR@CE, HUNT, BioVU, 23andMe, Gothenburg H70 Birth Cohort Studies and Clinical AD from Sweden (Gothenburg), ANMerge).

Proxy vs case-control LOAD

The genetic correlation between the proxy LOAD GWAS results and the case-control LOAD results was 0.83 (SE=0.21, $P=6.61x10^{-5}$), which is on par with the genetic correlation between proxy and case-control LOAD in our previous study[1]. The high correlation suggests that the proxy design is a good estimate for LOAD status when the genotyped individual is too young to present the phenotype. However, there are differences between the results when specifying the phenotypes differently. Supplementary Figure 1 shows that the previously unidentified regions identified in the full meta-analysis do not have much significance in the proxy data alone. The *TMEM106B*, *GRN*, and *NTN5* regions did not have any variants with a P-value <0.0005 so none of the variants in that region were included in the Manhattan plot. Interestingly, *TMEM106B* and *GRN* are genes previously associated with frontotemporal dementia[2] and one would expect the LOAD proxy results to be driving this association due to the inclusion of non-medically diagnosed individuals but the association signal appears to be absent in the proxy results. Supplementary Figure 2 shows that these genes do have relatively strong, albeit non-significant, signals in the results from the case-control data. Further exploration of the previously unidentified regions in an independent sample will be valuable in determining the role of these genes in LOAD.

Genetic correlations

Across 855 external phenotypes in LDhub[3], two significant genetic correlations with the meta-analysis results were observed (**Supplementary Table 6**). The strongest correlation was with a previous LOAD study conducted by Lambert *et al*. (2013)[4] ($r_g=1.18$, SE=0.19, $P_{Bonferroni}=2.42x10^{-7}$). The other significant correlation was with the UKB[5] trait "Illnesses of mother: Alzheimer's disease/dementia" ($r_g=0.80$, SE=0.11, $P_{Bonferroni}=6.38x10^{-10}$). The current study included individuals which were also included in Lambert *et al*. (2013)[4] and the UKB.

Genomic risk loci enrichment

Active chromatin enrichment

The genomic risk loci (excluding the *HLA-DRB1* (MHC) region) contained 45,479 variants in total. An insight into the functional annotation of these variants may highlight potential routes from variant to phenotype. All the variants in the genome were annotated as being in active or inactive chromatin across 127 cell types based on the ROADMAP Core 15-state model[6] In all 127 cell types, the genomic risk loci variants were significantly enriched in variants within active chromatin compared to all variants included in the meta-analysis (**Supplementary Table 13**). The odds ratio (OR) of enrichment ranged from 4.34 to 1.71, with the top 5 cell types consisting of immune related cell types (**Supplementary Figure 16**). The least enriched cell types were also significantly enriched compared to the rest of the genome, this is likely due to the genomic risk loci being located in gene dense regions which are likely to be more active in all cell types compared to the whole genome. The pattern of enrichment and proportion of active chromatin across the cell types prioritizes immune cells as cell types of interest.

Functional consequence enrichment

The variants within and outside the genomic risk loci were also compared based on their functional consequence determined by ANNOVAR[7] (**Supplementary Figure 17; Supplementary Table 14**). The majority of the variants within the genomic risk loci are intronic and intergenic (prop=0.49, prop=0.36). The intergenic and ncRNA intronic variants were the only variant types to be significantly depleted in the genomic risk loci (OR=0.72, $P_{bonferroni}$=<1x10^{-323}; OR=0.76, $P_{bonferroni}$=1.20x10^{-39}), all other annotations, except ncRNA splicing variants, were significantly enriched. Splicing variants were the most enriched (OR=4.16, $P_{bonferroni}$=0.0098). These results suggest that the genomic risk loci are regions rich in genes, and that splicing may be an important mechanism through which effects of these genes on LOAD are regulated.

Gene prioritization

Previously unidentified loci

The lead variant in locus 1 is a rare (MAF=0.0041) synonymous variant (rs113020870) located in *AGRN* (**Supplementary Figure 18**). The lead variant is supported by only a handful of other variants in LD (R^2>0.1), one of which is near significance (rs575531402, P=2.24x10^{-7}). Neither of these variants are known GTEx eQTLs. The lead variant is present in 3 datasets (23andMe, Finngen, and UKB). The effect direction in all 3 is the same and the *P*-values in the UKB and 23andMe datasets are below 5x10^{-4}. *AGRN* codes for agrin, a protein which influences the functioning of excitatory synapse and the blood brain barrier and has been previously suggested to be important to neurological diseases like LOAD[8]. Another study identified that *AGRN* expression influenced amyloid-beta homeostasis in mouse models of LOAD[9]. The follow-up analyses from our study does not strongly support *AGRN* as a LOAD causal gene, however previous literature provides some support that *AGRN* could be a causal gene for this locus.

The lead variant (rs2452170, P=1.72x10^{-8}) of locus 34 is located between *FUT2* and *MAMSTR* (**Supplementary Figure 19**). This region maps to 16 genes, three of which include exonic variants with high CADD scores (>20). Of those mapped genes, *CARD8*, *FAM83E*, *CA11*, *SEC1P*, *NTN5*, and *FUT2* are significantly differently expressed in bulk brain tissue of LOAD patients compared to controls[10]. *FUT2* contains rs601338, a variant

with a CADD score of 52 which introduces a stop codon. *FUT2* also contains another stop codon-introducing variant (rs602662), which has a CADD score of 22.5. The former variant has been previously associated with inflammatory bowel disease and Crohn's disease[11], and the latter was associated with B12 blood levels[12]. Another exonic variant (rs2287922) with a high CADD score (26.2) is located in *RASIP1*. *RASIP1* appears to be involved with maintenance of the blood brain barrier[13] and blood brain barrier malfunction has been observed in AD[14]. In a recent unpublished study, *MAMSTR* was identified as a gene which is differentially methylated in LOAD brains[15]. *NTN5* was implicated through colocalization of the GWAS signal with an eQTL (rs2452170) in brain tissue (ROSMAP). The identified eQTL is the lead variant and also an eQTL for *FUT2* and *MAMSTR* in GTEx brain tissue. *NTN5* is highly expressed in neurogenic regions of the brain and is known to be involved in adult neurogenesis[16]. It is difficult to prioritize a gene at this locus because there is evidence from this study and previous literature for *FUT2*, *MAMSTR*, *RASIP1*, and *NTN5*. There is ~3.8Mb of distance between the lead variants of the *APOE* region and locus 34 (R^2<0.001 in 1KG Europeans); however, the strength of the *APOE* region could be influencing this region so the interpretation of this locus is limited until further statistical and experimental analysis can identify the independence of locus 34.

Loci with high confidence (PIP >0.95) fine-mapped variants

Fine-mapping in locus 2 identified the lead variant (rs4663105) as the most likely causal variant (PIP=1). This variant is upstream of *BIN1* and is an eQTL in GTEx brain tissue. *BIN1* is a known LOAD gene and is thought to modulate tau pathology[17]. Fine-mapping in locus 10 identified the lead variant (rs187370608) and rs143332484 as causal variants (PIP>0.98). The latter of which is a *TREM2* missense variant with a CADD score of 11.8. *TREM2* is a known LOAD gene and has a role in microglia phagocytosis, and inflammation[18]. Fine-mapping in locus 16 identified a *SPATC1* intronic variant (rs79832570) as a causal variant (PIP=0.97). However, we were unable to identify previous literature to support the role of this variant and *SPATC1* in LOAD pathology. Another significant variant (rs34674752; P=5.54x10^{-9}) in the region is a *SHARPIN* missense variant with a CADD score of 21.9. A previous study identified a rare variant in *SHARPIN* associated with inflammation and LOAD[19]. Fine-mapping in locus 22 highlighted the lead variant (rs11218343) as a high probability variant (PIP=1). The lead variant is a low CADD score (3.60) intron variant in *SORL1*. *SORL1* is highly expressed in microglia in the Allen Human Brain Atlas[20]. *SORL1* is a known LOAD gene and has been implicated in *APP* trafficking and amyloid build up[21]. Fine-mapping in locus 32 identified the lead variant (rs12151021) as a high probability causal variant (PIP=1). The lead variant is a low CADD score (3.09) intron variant in *ABCA7*. There is some previous evidence for *ABCA7* in previous literature[22]. Five out of seven of the high PIP variants are the lead variants in their locus which suggests that the fine-mapping is not adding much additional information for gene prioritization. The two non-lead high PIP variants highlighted TREM2, a known LOAD gene, and *SPATC1*, a gene unlikely to be the causal gene. The fine-mapping analysis was limited by the use of an external reference panel and these results exemplify the importance of using very accurate LD information for fine-mapping.

Loci which colocalized with eQTLs

The signal in locus 19 colocalized with an eQTL for *MADD* in monocytes (BLUEPRINT). The lead variant (rs3740688) was the colocalized variant. The lead variant is a low CADD (0.37) intronic variant in *SPI1* and an eQTL for 19 genes in GTEx tissues. The large spread of association signal and correlated gene expression makes prioritizing genes in this locus difficult. Colocalization implicated the lead variant (rs117618017) as the colocalized variant in locus 26. The lead variant colocalized with *APH1B* expression in immune cells and whole blood, which was also observed in Schwartzentruber *et al.* (2021)[23].

The lead variant is an *APH1B* missense variant (CADD=10.57). The protein (aph-1 homolog B, gamma-secretase subunit) encoded by *APH1B* is part of the γ-secretase complex and has been implicated in *APP* processing[24]. The association signal in locus 30 colocalized with an eQTL for *AC004687.2* in Th2 memory cells (Schmiedel). The implicated variant (rs2526377) is an upstream variant of MIR142 and an intron variant of TSPOAP1-AS1. MIR142 is a microRNA that is predicted to target genes related to inflammation and neurodegeneration in the brain[25]. TSPOAP1-AS1 has also been reported as a LOAD gene and is a marker of neuroinflammation[26]. Prioritizing a gene in this locus is difficult because prior evidence is sparse and colocalization identified a poorly understood gene (*AC004687.2*). In locus 31, the lead variant (rs6504163) was implicated through colocalization with an eQTL for *ACE* in brain (ROSMAP and BrainSeq) and monocyte (Schmeidel). The lead variant is upstream of *ACE* and is an eQTL in GTEx brain tissue. Deletion of *ACE* in LOAD mouse models has been shown to increase amyloid deposition[27]. An upstream *CD33* variant (rs3865444) was implicated in locus 35 for colocalization with an eQTL for *CD33* in monocytes (BLUEPRINT) and lymphoblastoid cell lines (TwinsUK). *CD33* is thought to affect microglia activation and amyloid clearance[28]. *CASS4* was implicated in locus 37 by colocalization with eQTLs in monocytes (BLUEPRINT). The lead variant (rs6069737) was identified as the candidate for colocalization. The lead variant is a low CADD score (8.04) intronic variant in *CASS4*. *CASS4* codes for scaffolding proteins and it has been speculated that these proteins interact with amyloid and tau transport[29] but the role of *CASS4* in LOAD is not well characterized[30].

Brain regional gene expression

The regional brain expression of the genes implicated by eQTL mapping was examined using GAMBA[31] (**Supplementary Figure 20**). The mean gene expression of the 329 mapped genes were compared to a random selection of 329 significantly brain expressed genes which resulted in 27 regions which differed significantly (**Supplementary Table 15**). The significantly different regions with the highest mean were the left thalamus proper, caudal anterior cingulate, insula, pallidum, and postcentral gyrus. These regions have a range of functions including somatosensory function, emotion, and memory[32–36]. However, when the regional mean gene expressions of the mapped genes were compared to a random selection of genes (not just brain expressed genes) there were no significant differences between the mean gene expression of any region (**Supplementary Table 15**). This result may reflect the lack of brain specificity of the mapped genes and highlights the importance of narrowing down associated regions to individual causal genes.

Polygenic risk score

We generated polygenic risk scores (PRS) to identify how well the meta-analysis results could predict LOAD status using meta-analysis results from a subset of the data. The Gothenburg H70 Birth Cohort Studies and Clinical AD from Sweden (Gothenburg) dataset (712 cases, 2523 controls) was used as the independent test set because it was the dataset with the largest number of cases of the available genotype-level data. To identify effect estimates and standard errors for the PRS analysis, we performed a METAL[37] inverse-variance weighted meta-analysis, this meta-analysis excluded the independent test set (Gothenburg) and the UKB data. The UKB data was excluded from the METAL meta-analysis because the effect estimates of this dataset were calculated for a quantitative phenotype and were, therefore, not directly comparable to the other (case-control) datasets. The best model of the PRS analysis from the METAL meta-analysis results (43,013 cases and 715,456 controls) explained 5.3% of the variance in the Gothenburg data (Threshold=$5x10^{-8}$; P=$1.92x10^{-30}$) and 1.2% variance in the Gothenburg data (Threshold=$5x10^{-8}$; P=$1.83x10^{-8}$) when the larger *APOE* region (GRCh37: 19:40000000-

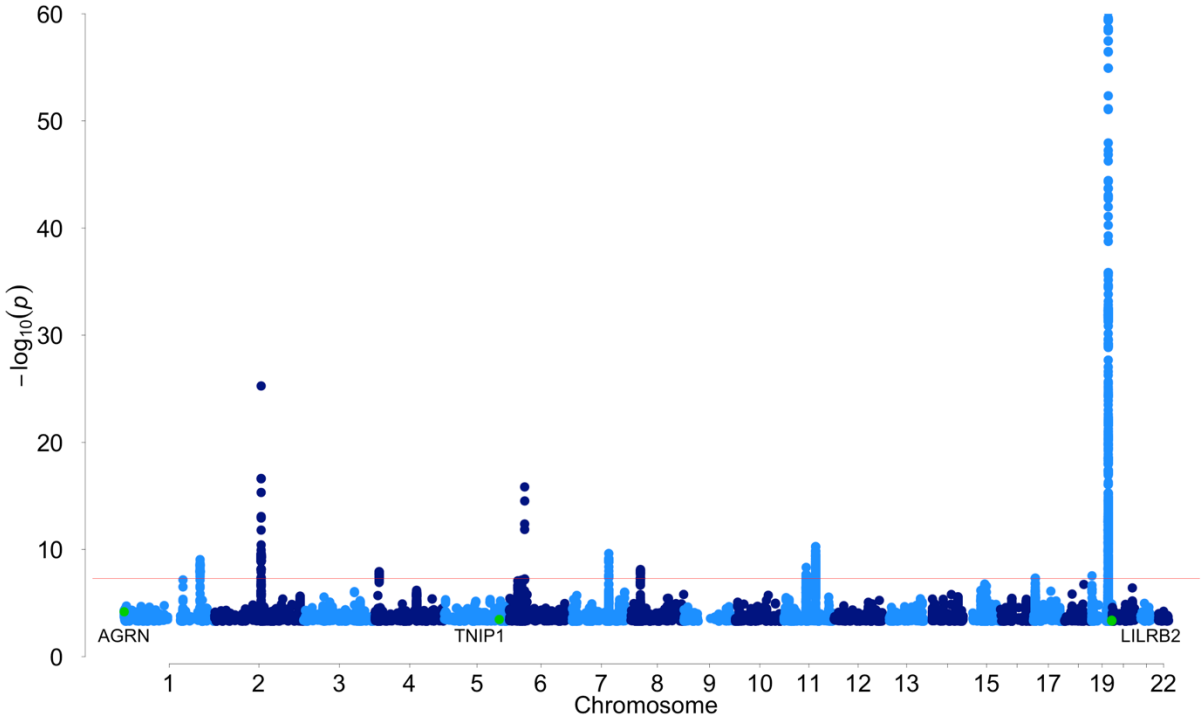
50000000) was excluded from the model. These results are similar to the estimates identified in our previous meta-analysis[1]. We performed the same PRS using the UKB summary statistics and found similar performance to the case-control data with (R^2= 0.066, Threshold=$5x10^{-8}$; P=$2.49x10^{-37}$) and without the larger *APOE* region (R^2= 0.009, Threshold=$5x10^{-8}$; P=$2.49x10^{-37}$). The base model (without covariates) PRS of the case-control and UKB models were included in a logistic regression where the combination of both PRS models explained slightly more variance than either alone (R^2=0.075).

Enrichment analyses using replicated loci

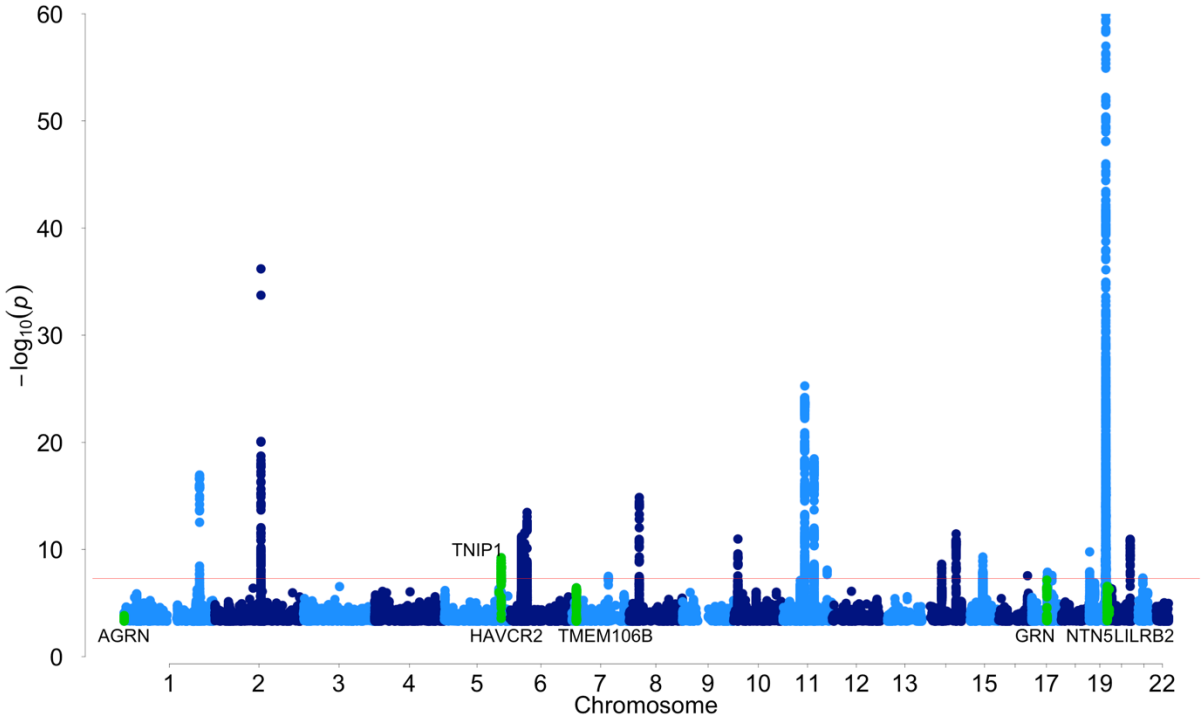
Functional follow-up restricted to the significant regions which were previously identified in Kunkle *et al.* (2019)[38] found similar results to the functional follow-up using the entire meta-analysis results (**Supplementary Table 17**). When significant loci which did not occur in Kunkle *et al.* (2019) were removed, spleen was no longer a significant associated tissue ($P_{Bonferroni}$= 0.16), one microglia dataset was no longer significantly associated (GSE104276_Human_Prefrontal_cortex_all_ages), and 4 gene-sets were no longer significantly associated (GO_bp:go_amyloid_precursor_protein_metabolic_process, GO_bp:go_positive_regulation_of_amyloid_beta_formation, GO_bp:go_immune_response_inhibiting_signal_transduction, GO_bp:go_positive_regulation_of_amyloid_precursor_protein_catabolic_process). The removal of significant loci which did not occur in Kunkle *et al.* (2019) caused the gain of one significant gene-set (GO_mf:go_apolipoprotein_binding). The genomic risk loci are significantly enriched for active chromatin in the same cell types in the full analysis and the analyses of only significant loci which occur in Kunkle *et al.* (2019). The same functional consequences are significantly associated with the genomic risk loci, except non-coding RNA exonic (ncRNA_exonic) variants, in the full analysis and the analyses of only significant loci which occur in Kunkle *et al.* (2019). Largely, this demonstrates that many of the enrichment findings are robust to the exclusion of significant loci which have only been identified in this study and previous meta-analysis which include proxy datasets. It also demonstrates that the addition of the loci identified in this study provide extra information on a genome wide and individual locus level.

Supplementary Figures

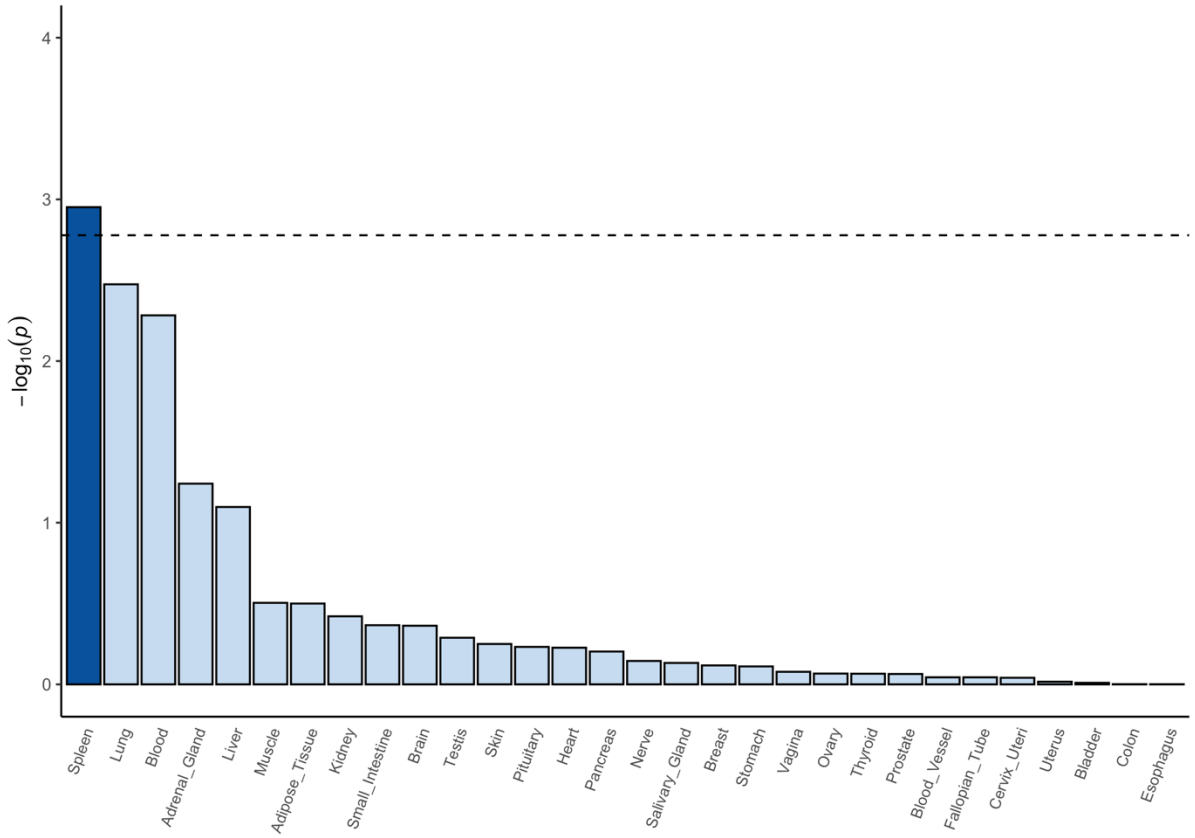
Supplementary Figure 1: Manhattan plot of the UKB proxy LOAD data indicates low association of the previously unidentified regions (green) identified in the full meta-analysis. Only variants with a *P*< 0.0005 are displayed so previously unidentified regions with P-values larger than this are not observable. The P-values were identified through linear regression (two-sided test) and were not adjusted for multiple testing. *AGRN*, *TNIP1*, and *LILRB2* are the only previously unidentified regions with P-values small enough to be displayed. The *APOE* region cannot be fully observed because the y-axis is limited to the top variant in the second most significant locus, -log10(1x10^{-60}), in order to display the less significant variants. The red line represents genome wide significance (5x10^{-8}).



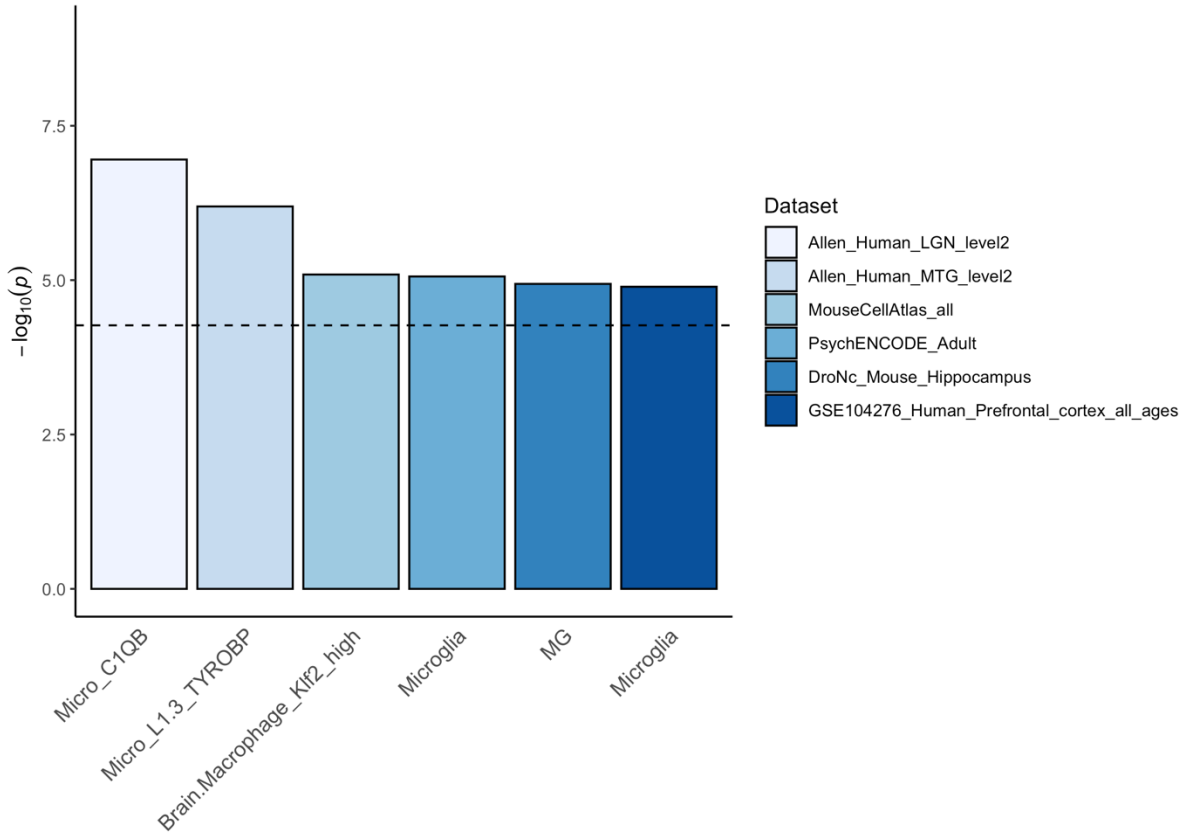
Supplementary Figure 2: Manhattan plot of the meta-analysis results of the LOAD case-control data (UKB data excluded) indicates the association of the previously unidentified loci (green) identified in the full meta-analysis. Only variants with a *P*< 0.0005 are displayed. The P-values were identified through a meta-analysis (two-sided test) of summary statistics generated by logistic regressions (two-sided test) and were not adjusted for multiple testing. The *APOE* region cannot be fully observed because the y-axis is limited to the top variant in the second most significant locus, -log10($1x10^{-60}$), in order to display the less significant variants. The red line represents genome wide significance ($5x10^{-8}$). The *TNIP1*/*HAVCR2* regions and the *NTN5*/*LILRB2* regions are close enough together that they cannot be visually distinguished at this scale but are different genomic risk loci.



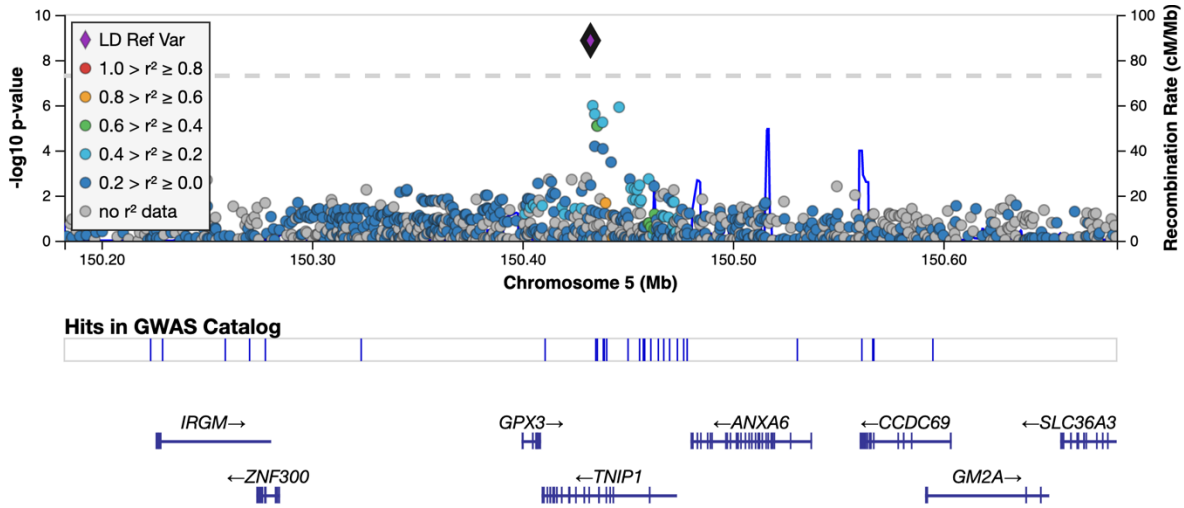
Supplementary Figure 3: MAGMA tissue specificity analysis identified spleen as the only GTEx tissue where expression of the MAGMA genes was significantly associated (one-sided test). The dotted line represents the significance threshold after Bonferroni correction for 30 tests. The significantly associated tissue is highlighted in dark blue. The full results are available in Supplementary Table 6.



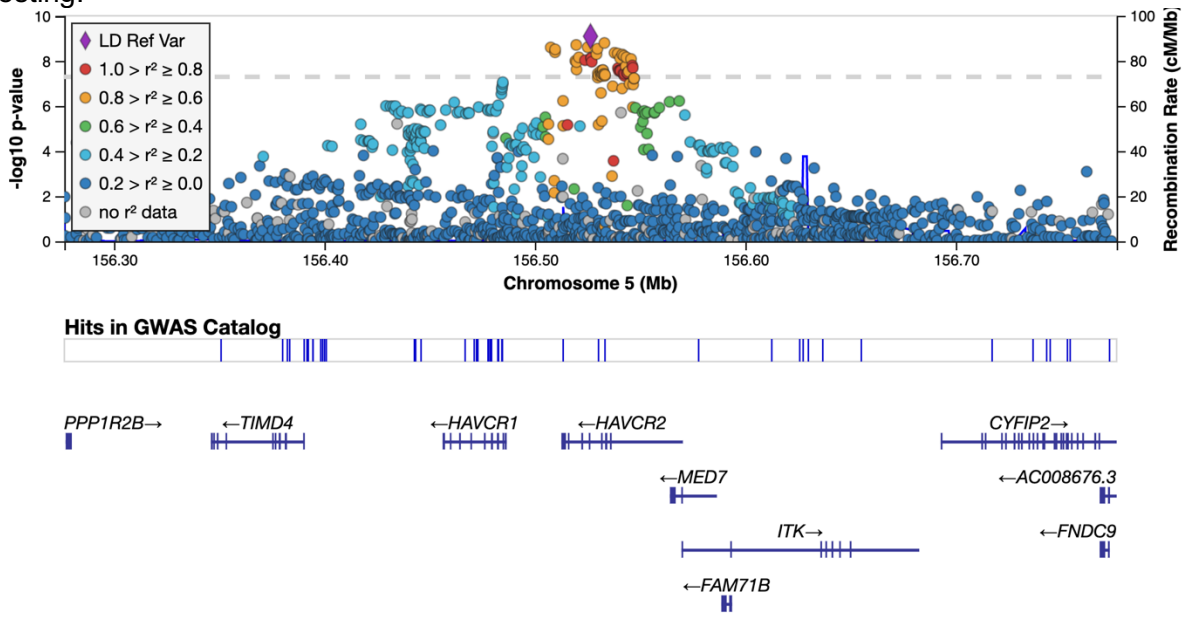
Supplementary Figure 4: Independent cell type associations based on within-dataset corrected analyses (one-sided test) identifies microglia as the only cell type of interest. Six single-cell datasets were significantly associated, after Bonferroni correction, with the expression of LOAD-associated genes. The dotted line represents the significance threshold after Bonferroni correction (P<5.39x10^{-5}). Microglia were significantly associated in human lateral geniculate nucleus (P=1.11x10^{-7}), human middle temporal gyrus (P=6.41x10^{-7}), adult human brain (P=8.72x10^{-6}), mouse hippocampus (P=1.15x10^{-5}), human prefrontal cortex (P=1.28x10^{-5}), and brain macrophage (microglia) mouse brain (P=8.11x10^{-6}).



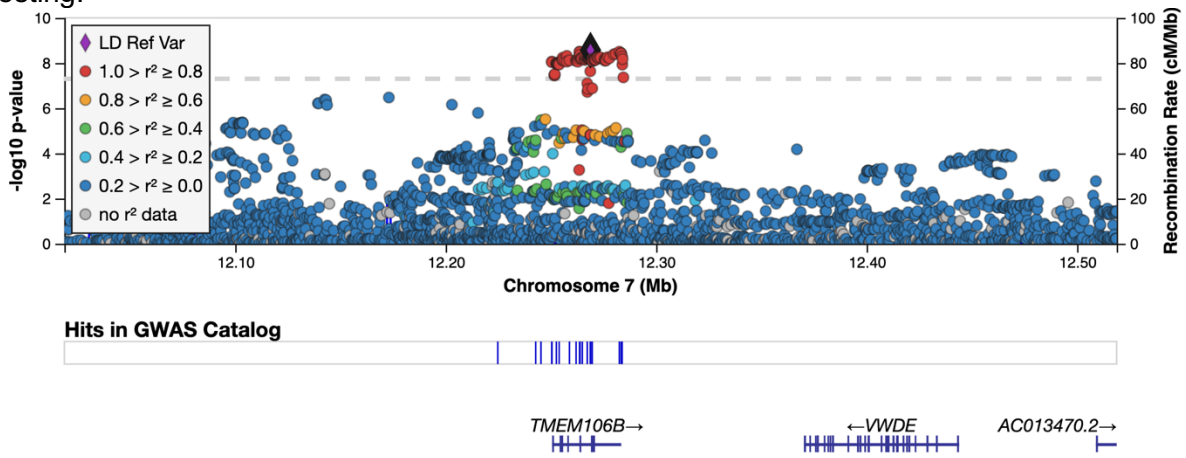
Supplementary Figure 5: Regional plot highlighting the lead variant of locus 7 and the nearby genes. There were no variants in LD with the lead variant in the 1KG European population so the LD estimates were obtained from the 1KG East Asian population to highlight variants in LD with the lead variant. The P-values were identified through a meta-analysis (two-sided test) of summary statistics generated by linear/logistic regressions (two-sided test) and were not adjusted for multiple testing.



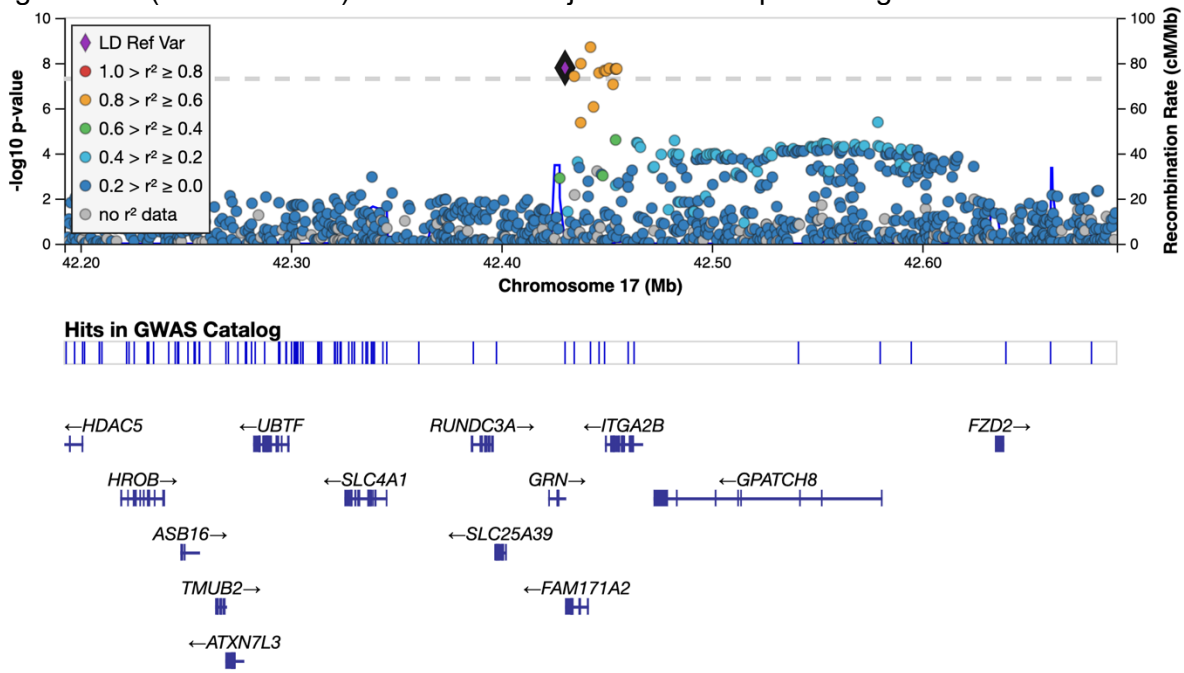
Supplementary Figure 6: Regional plot highlighting the lead variant of locus 8 and the nearby genes. The LD information was obtained from the 1KG European population. The P-values were identified through a meta-analysis (two-sided test) of summary statistics generated by linear/logistic regressions (two-sided test) and were not adjusted for multiple testing.



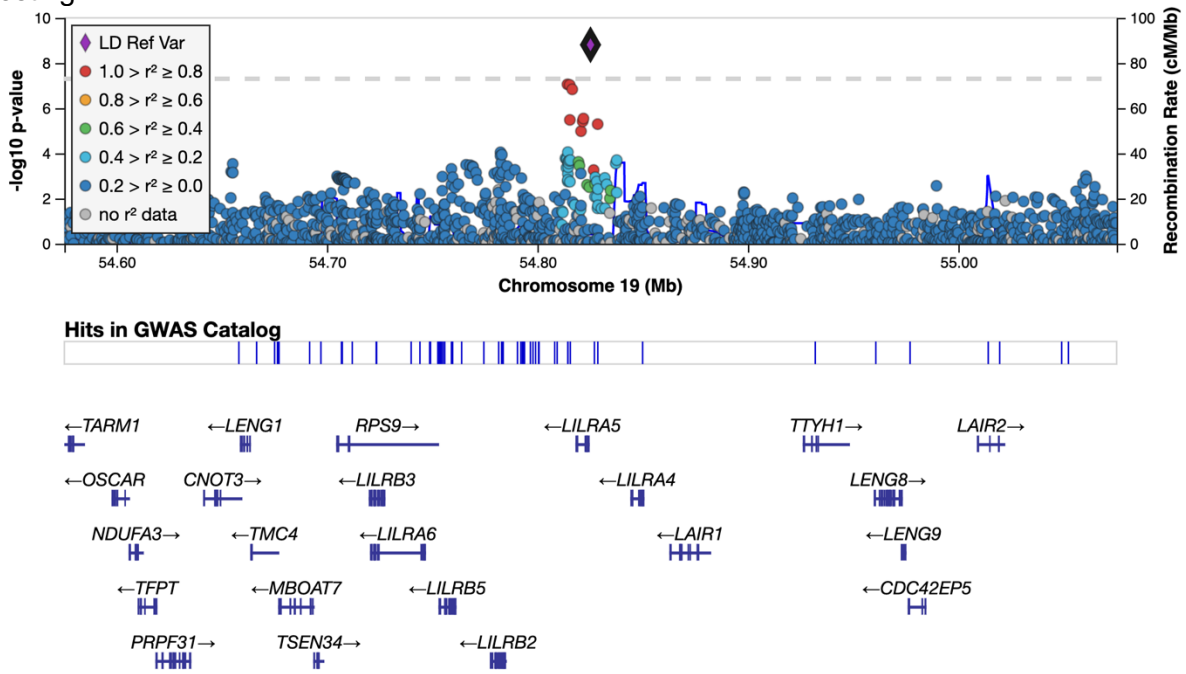
Supplementary Figure 7: Regional plot highlighting the lead variant of locus 12 and the nearby genes. The LD information was obtained from the 1KG European population. The P-values were identified through a meta-analysis (two-sided test) of summary statistics generated by linear/logistic regressions (two-sided test) and were not adjusted for multiple testing.



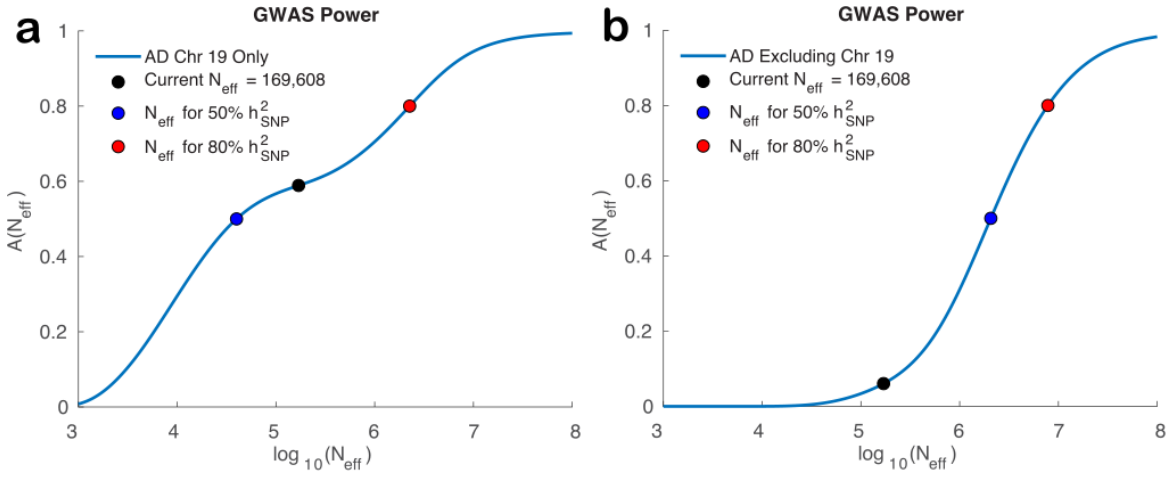
Supplementary Figure 8: Regional plot highlighting the lead variant of locus 28 and the nearby genes. The LD information was obtained from the 1KG European population. There was no LD information for the lead variant so the nearest significant variant with LD information was chosen as the LD reference variant (rs5848). The P-values were identified through a meta-analysis (two-sided test) of summary statistics generated by linear/logistic regressions (two-sided test) and were not adjusted for multiple testing.



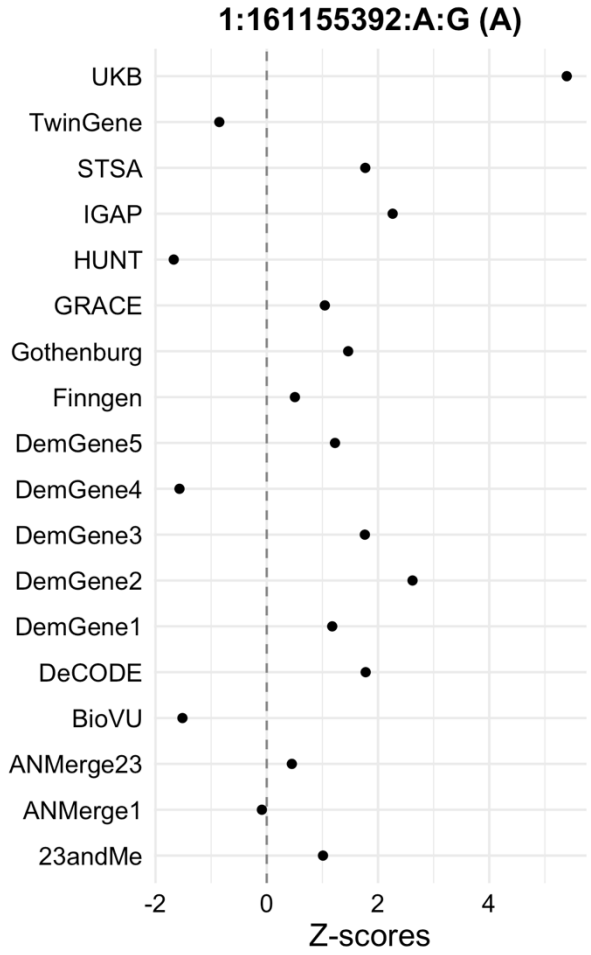
Supplementary Figure 9: Regional plot highlighting the lead variant of locus 36 and the nearby genes. The LD information was obtained from the 1KG European population. The P-values were identified through a meta-analysis (two-sided test) of summary statistics generated by linear/logistic regressions (two-sided test) and were not adjusted for multiple testing.



Supplementary Figure 10: The predicted power of our meta-analysis to explain the total heritability from chromosome 19 (**a**) and all other autosomes (**b**). The current sample size estimate is indicated by the black point. $A(N_{eff})$ represents the proportion of SNP heritability explained by the effective sample size. $Log_{10}(N_{eff})$ represents the log10 effective sample size.

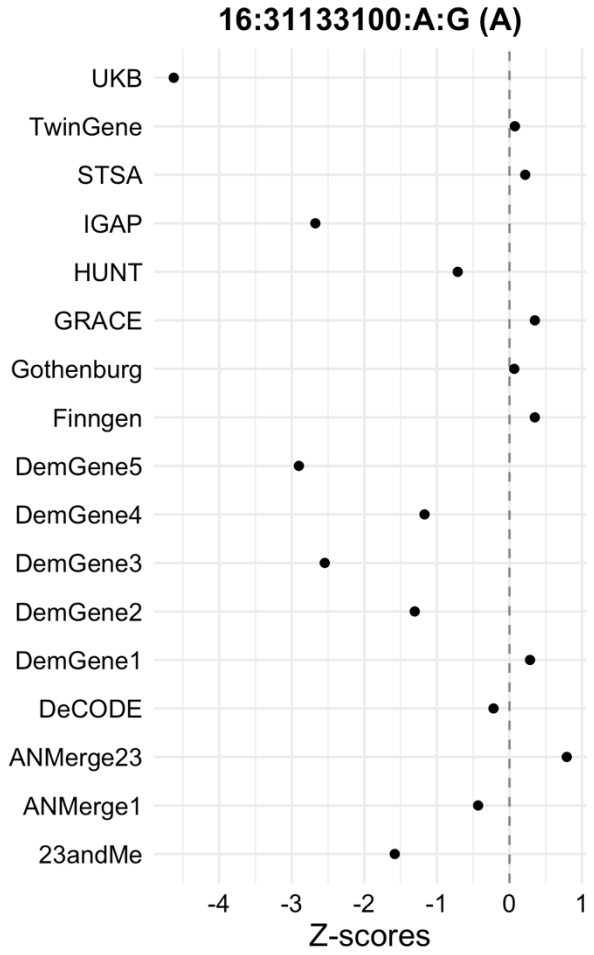


Supplementary Figure 11: The per dataset Z-scores and P-values of the lead variant (rs4575098; 1:161155392:A:G) of the *ADAMTS4* region in Jansen *et al.* (2019)[1] outlines the contribution of each dataset to the non-significant result in the meta-analysis of this study. The Z-scores are aligned to the allele in brackets. The summary statistics were generated by linear/logistic regressions (two-sided test) and were not adjusted for multiple testing. The points represent the Z-scores in each dataset.



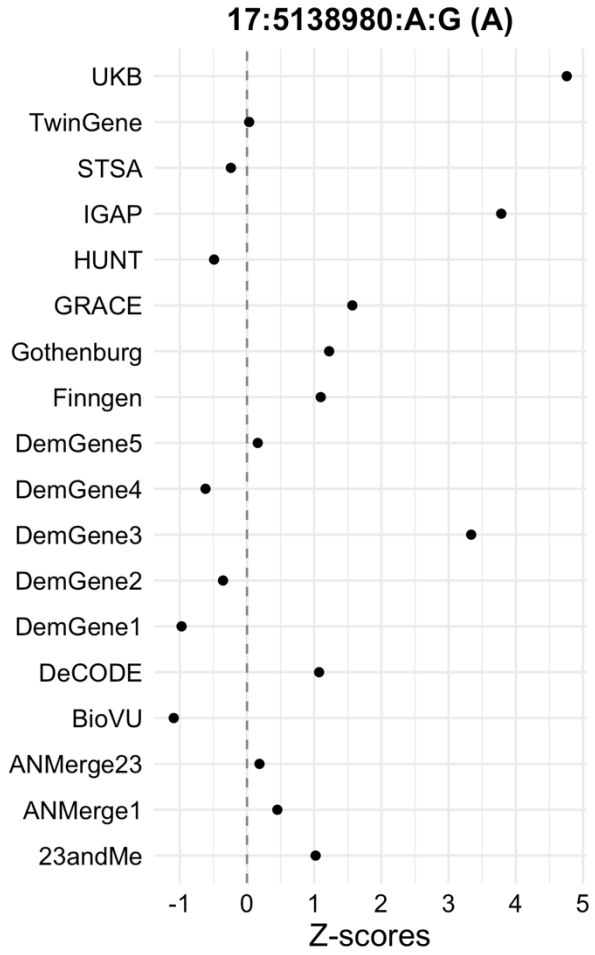
Dataset	Z	P
UKB	5.3900	6.88e-08
TwinGene	-0.8530	3.93e-01
STSA	1.7700	7.71e-02
IGAP	2.2600	2.34e-02
HUNT	-1.6700	9.45e-02
Gothenburg	1.4600	1.44e-01
GRACE	1.0400	2.96e-01
Finngen	0.5060	6.13e-01
DemGene5	1.2300	2.20e-01
DemGene4	-1.5700	1.17e-01
DemGene3	1.7600	7.84e-02
DemGene2	2.6200	8.68e-03
DemGene1	1.1800	2.39e-01
DeCODE	1.7800	7.54e-02
BioVU	-1.5200	1.30e-01
ANMerge23	0.4510	6.53e-01
ANMerge1	-0.0879	9.30e-01
23andMe	1.0100	3.13e-01
META-ANALYSIS	5.0600	4.22e-07

Supplementary Figure 12: The per dataset Z-scores and P-values of the lead variant (rs59735493; 16:31133100:A:G) of the *KAT8* region in Jansen *et al.* (2019)[1] outlines the contribution of each dataset to the non-significant result in the meta-analysis of this study. The Z-scores are aligned to the allele in brackets. The summary statistics were generated by linear/logistic regressions (two-sided test) and were not adjusted for multiple testing. The points represent the Z-scores in each dataset.



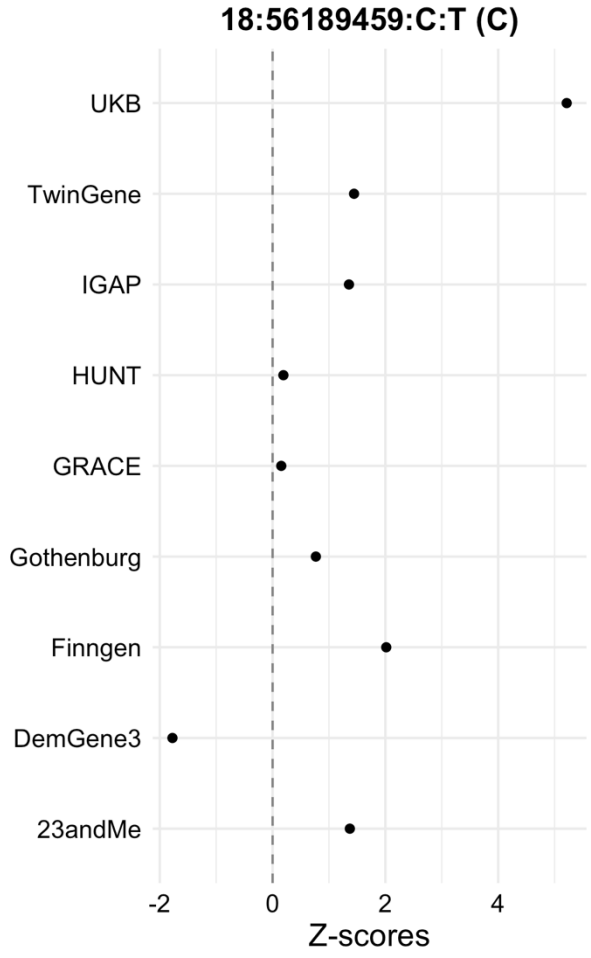
Dataset	Z	P
UKB	-4.6300	3.72e-06
TwinGene	0.0747	9.40e-01
STSA	0.2170	8.27e-01
IGAP	-2.6800	7.42e-03
HUNT	-0.7130	4.76e-01
Gothenburg	0.0660	9.45e-01
GRACE	0.3500	7.27e-01
Finngen	0.3500	7.26e-01
DemGene5	-2.9000	3.71e-03
DemGene4	-1.1700	2.42e-01
DemGene3	-2.5500	1.09e-02
DemGene2	-1.3100	1.92e-01
DemGene1	0.2830	7.77e-01
DeCODE	-0.2200	8.26e-01
ANMerge23	0.7900	4.28e-01
ANMerge1	-0.4320	6.66e-01
23andMe	-1.5800	1.13e-01
META-ANALYSIS	-4.7300	2.25e-06

Supplementary Figure 13: The per dataset Z-scores and P-values of the lead variant (rs113260531; 17:5138980:A:G) of the *SCIMP* region in Jansen *et al.* (2019)[1] outlines the contribution of each dataset to the non-significant result in the meta-analysis of this study. The Z-scores are aligned to the allele in brackets. The summary statistics were generated by linear/logistic regressions (two-sided test) and were not adjusted for multiple testing. The points represent the Z-scores in each dataset.



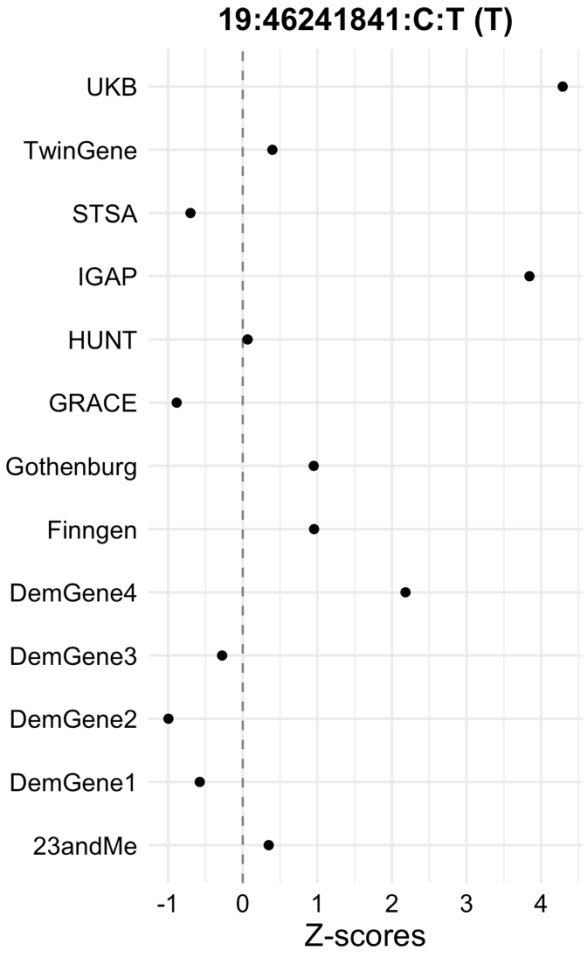
Dataset	Z	P
UKB	4.7600	1.94e-06
TwinGene	0.0327	9.75e-01
STSA	-0.2390	8.11e-01
IGAP	3.7800	1.56e-04
HUNT	-0.4890	6.25e-01
Gothenburg	1.2200	2.21e-01
GRACE	1.5700	1.17e-01
Finngen	1.1000	2.73e-01
DemGene5	0.1600	8.73e-01
DemGene4	-0.6170	5.37e-01
DemGene3	3.3400	8.52e-04
DemGene2	-0.3560	7.22e-01
DemGene1	-0.9730	3.30e-01
DeCODE	1.0700	2.83e-01
BioVU	-1.0900	2.76e-01
ANMerge23	0.1860	8.52e-01
ANMerge1	0.4520	6.52e-01
23andMe	1.0200	3.09e-01
META-ANALYSIS	5.3700	7.85e-08

Supplementary Figure 14: The per dataset Z-scores and P-values of the lead variant (rs76726049; 18:56189459:C:T) of the *ALPK2* region in Jansen *et al.* (2019)[1] outlines the contribution of each dataset to the non-significant result in the meta-analysis of this study. The Z-scores are aligned to the allele in brackets. The summary statistics were generated by linear/logistic regressions (two-sided test) and were not adjusted for multiple testing. The points represent the Z-scores in each dataset.



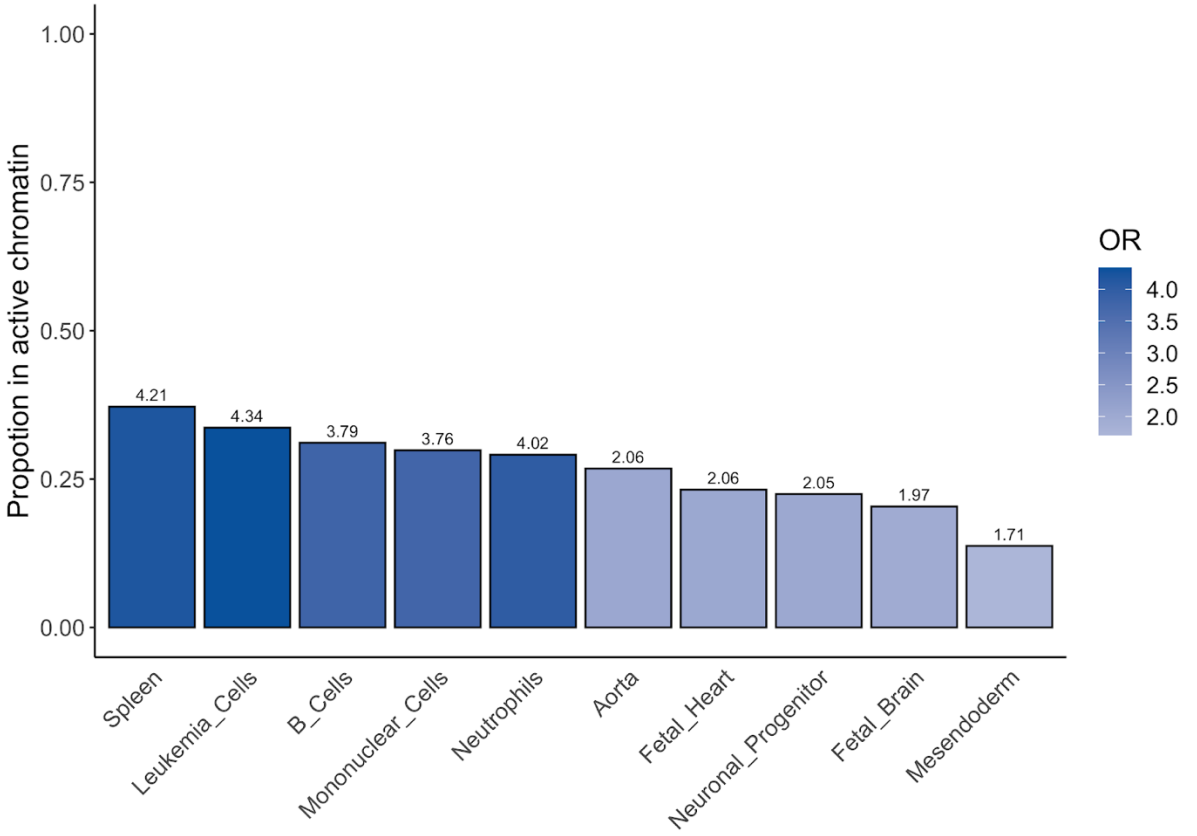
Dataset	Z	P
UKB	5.220	1.83e-07
TwinGene	1.440	1.49e-01
IGAP	1.350	1.76e-01
HUNT	0.192	8.48e-01
Gothenburg	0.767	4.43e-01
GRACE	0.154	8.78e-01
Finngen	2.020	4.38e-02
DemGene3	-1.780	7.58e-02
23andMe	1.370	1.79e-01
META-ANALYSIS	5.120	3.08e-07

Supplementary Figure 15: The per dataset Z-scores and P-values of the lead variant (rs76320948; 19:46241841:C:T) of the *AC074212.3* region in Jansen *et al.* (2019)[1] outlines the contribution of each dataset to the non-significant result in the meta-analysis of this study. The Z-scores are aligned to the allele in brackets. The summary statistics were generated by linear/logistic regressions (two-sided test) and were not adjusted for multiple testing. The points represent the Z-scores in each dataset.

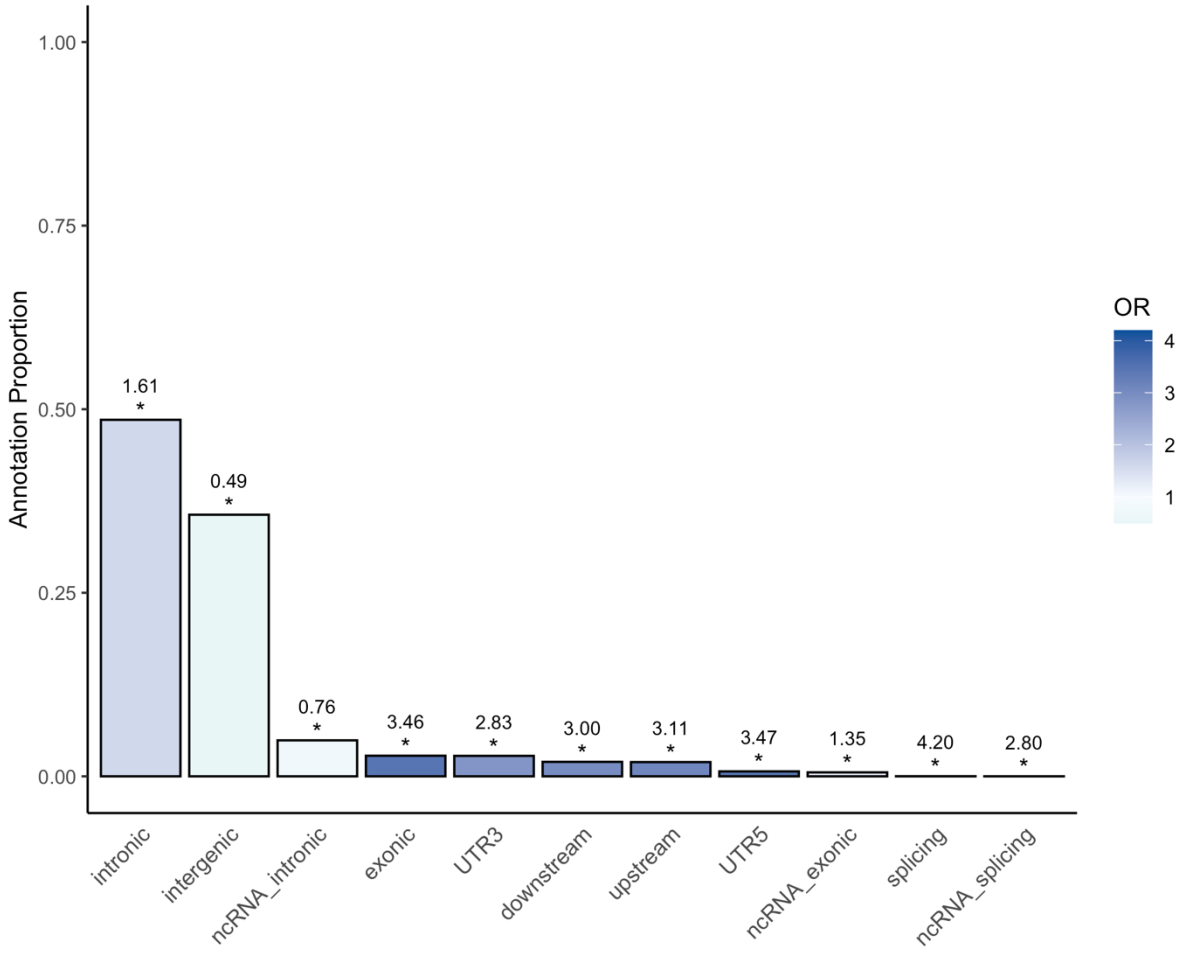


Dataset	Z	P
UKB	4.2900	1.80e-05
TwinGene	0.3980	6.92e-01
STSA	-0.7000	4.84e-01
IGAP	3.8400	1.22e-04
HUNT	0.0653	9.48e-01
Gothenburg	0.9510	3.42e-01
GRACE	-0.8850	3.76e-01
Finngen	0.9560	3.39e-01
DemGene4	2.1800	2.92e-02
DemGene3	-0.2770	7.82e-01
DemGene2	-0.9960	3.20e-01
DemGene1	-0.5760	5.65e-01
23andMe	0.3480	7.29e-01
META-ANALYSIS	4.1800	2.96e-05

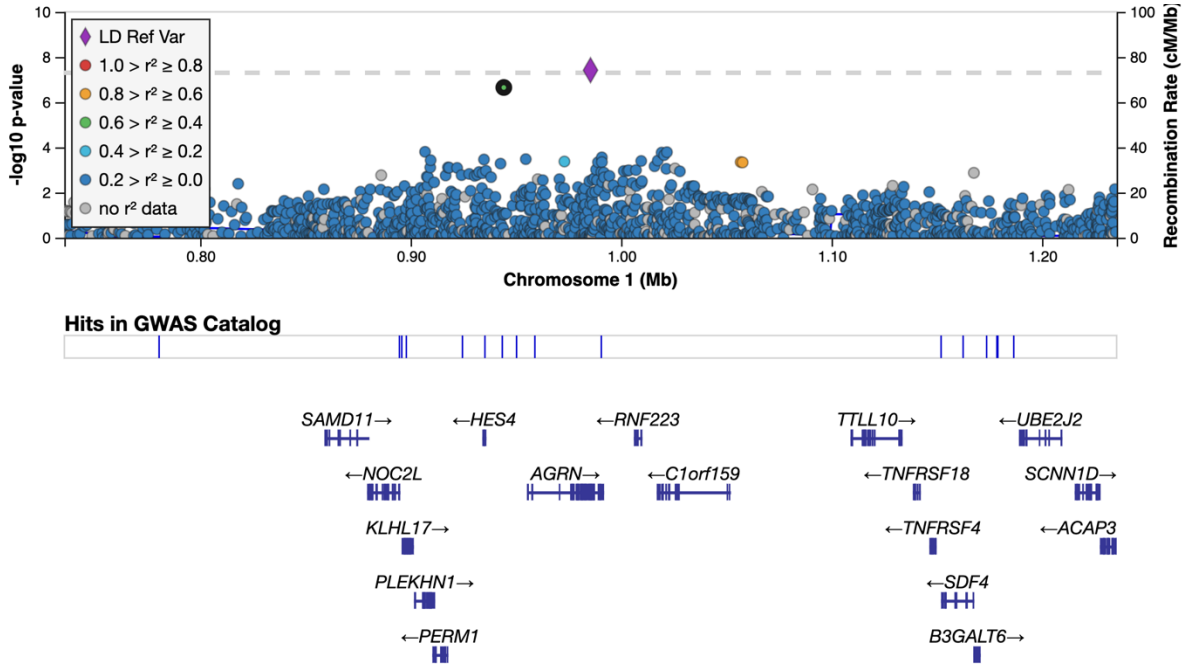
Supplementary Figure 16: The top 5 and bottom 5 cell types where active chromatin is most enriched in genomic risk loci compared to the rest of the genome highlights immune cells as enriched for active chromatin in LOAD regions of interest. The y-axis represents the proportion of the genomic risk loci that are in active chromatin in that cell type. The colour and labels of the bars represents the odds ratio (OR) from a Fisher's exact test (two-sided) comparing counts of variants in active chromatin in the genomic risk loci vs counts of variants in active chromatin in the rest of the genome. All OR were significantly different from 1 after Bonferroni correction.



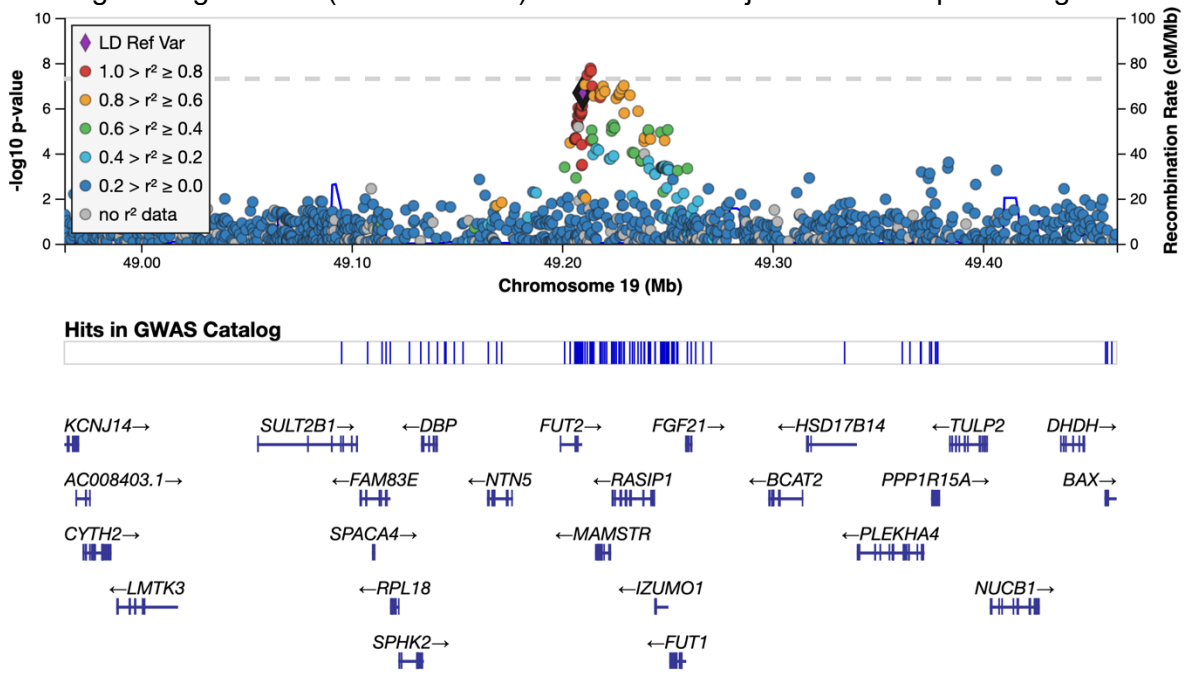
Supplementary Figure 17: ANNOVAR enrichment analyses identifies 10 significant differences between the number of annotations in the genomic risk loci compared to the rest of the genome. The y-axis represents the proportion of the genomic risk loci which falls into each annotation. The colour and the labels of the bars represents the odds ratio (OR) from a Fisher's exact test (two-sided) comparing counts of annotations in the genomic risk loci vs counts of annotations in the rest of the genome. The asterisks (*) represent OR which are significantly different from 1 after Bonferroni correction.



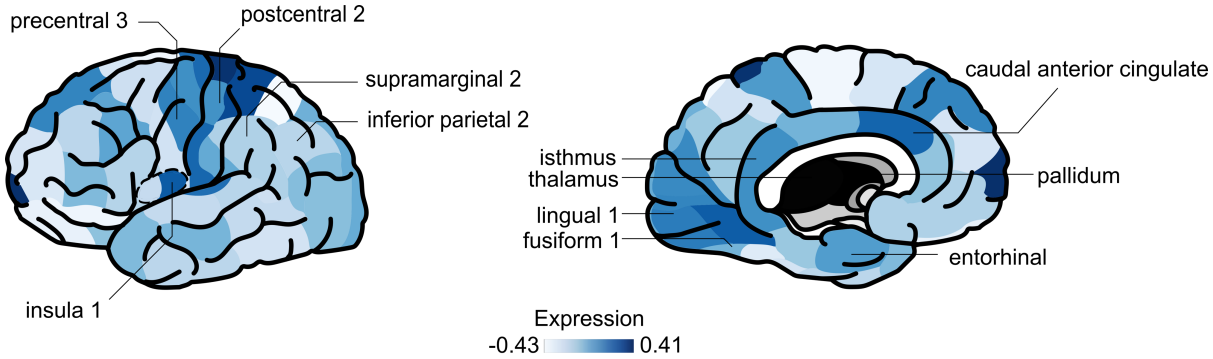
Supplementary Figure 18: Regional plot of locus 1 indicating the lead variant and the nearby genes. The LD estimates are derived from the 1KG European population. The P-values were identified through a meta-analysis (two-sided test) of summary statistics generated by linear/logistic regressions (two-sided test) and were not adjusted for multiple testing.



Supplementary Figure 19: Regional plot highlighting the lead variant of locus 34 and the nearby genes. The LD information was obtained from the 1KG European population. There was no LD information for the lead variant so the nearest suggestive variant with LD information was chosen as the LD reference variant (rs2638281). The P-values were identified through a meta-analysis (two-sided test) of summary statistics generated by linear/logistic regressions (two-sided test) and were not adjusted for multiple testing.



Supplementary Figure 20: The mean gene expression of 329 genes linked to risk loci through eQTL mapping across 64 brain regions highlights the top 10 regions with the highest expression. Gene expression values are obtained from the Allen Human Brain Atlas.



Supplementary Methods

Datasets

deCODE

Data from the deCODE study included 7,002 Alzheimer's patients (5,098 of whom were chip-typed) and 181,573 controls (88,739 of whom were chip-typed). In 15% of patients, the diagnosis of Alzheimer's disease was established at the Memory Clinic of the University Hospital according to the criteria for definite, probable, or possible Alzheimer's disease of the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA). In 80% of patients, the diagnosis has been registered according to the criteria for code 331.0 in ICD-9, or for F00 and G30 in ICD-10 in health records. Five percent of the patients were identified in the Directorate of Health medication database as having been prescribed Donepezil (Aricept). The controls were drawn from various research projects at deCODE Genetics. The study was approved by the National Bioethics Committee and the Icelandic Data Protection Authority. Written informed consent was obtained from all participants or their guardians before blood samples were drawn. All sample identifiers were encrypted in accordance with the regulations of the Icelandic Data Protection Authority.

Chip-typing and long-range phasing of 155,250 individuals was carried out as described previously[39]. Imputation of the variants found in 28,075 whole-genome sequenced individuals into the chip-typed individuals and 285,664 close relatives was performed as detailed earlier[39]. Association analysis in the deCODE sample was carried out using logistic regression with AD status as the response and genotype counts and a set of nuisance variables, including sex, county of birth, and current age, as predictors[40]. Correction for inflation of test statistics due to relatedness and population stratification in this Icelandic cohort was performed using the intercept estimate (1.30) from LD score regression[41].

UK Biobank

The UK Biobank (UKB; www.ukbiobank.ac.uk)[5] summary statistics for 46,613 cases and 318,246 controls were obtained from Jansen *et al.* (2019)[1]. In short, a proxy phenotype for Alzheimer's disease case-control status was generated from a self-report questionnaire which asked participants to report whether their biological mother or father ever suffered from Alzheimer's disease/dementia, and to report each parent's current age (or age at death, if applicable). The phenotype was constructed as a count of the number of affected parents ranging from 0 to 2. The contribution for each unaffected parent to the phenotype was weighted by the parent's age/age at death. This was calculated as the ratio of parent's age to age 100 (weight=(100-age)/100). The weight for an unaffected parent was capped at 0.32, corresponding to a risk equivalent to that of the maximum population prevalence of AD. Participants with a diagnosis of "Alzheimer's disease" (code G30) or "Dementia in Alzheimer's disease" (code F00) were given the maximum possible score of 2. Standard QC procedures were applied to the genotype data which was then imputed to the HRC[42], 1KG[43], and UK10K reference panels[44]. Further information on the quality control is available in Jansen *et al.* (2019)[1]. The imputed data was analyzed using linear regression with 12 ancestry principal components, age, sex, genotyping array, and assessment centre included as covariates. All participants provided written informed consent; the UKB received ethical approval from the National Research Ethics Service Committee North West-Haydock (reference 11/NW/0382), and all study procedures were in accordance with the World

Medical Association for medical research. Access to the UK Biobank data was obtained under application number 16406.

The Trøndelag Health Study (HUNT)

The HUNT data consists of 1156 cases and 7157 controls, where cases were defined as individuals diagnosed with ICD-10 G30.0 or F00*, or ICD-9 331.0 and controls were individuals last seen as healthy with no previous diagnosis of Alzheimer's disease. All controls were more than 80 years old. Participants overlapping with the DemGene study were removed. Further information about the biobank is available at https://hunt-db.medisin.ntnu.no/hunt-db/#/.

The samples were genotyped with Illumina HumanCoreExome arrays (HumanCoreExome12 v1.0, HumanCoreExome12 v1.1, or UM HUNT Biobank v1.0). Participants with call rates <99%, contamination >2.5%, large chromosomal copy number variants, lower call rate of technical duplicate pair or twins, uncommon sex chromosomal conformations, or discrepancies in reported gender were removed. The samples passing QC were analysed in a second round of genotype calling, using the Genome Studio quality control described elsewhere[45]. Variants were mapped to the Genome Reference Consortium Human genome build 37 (http://genome.ucsc.edu) using BLAT[46].

Variants were excluded if they had call rates <99%, higher call rates in another assay, probe sequences not mapping to the reference genome, cluster separation <0.3, gentrain score <0.15, or Hardy Weinberg equilibrium deviation (*P-value*<0.0001) from unrelated European samples. We also removed variants with frequency differences > 15% between the datasets or that were monomorphic in one dataset and had MAF > 1% in one of the others. Only European ancestry individuals were included. Ancestry was inferred using PLINKv1.90[47], projecting the HUNT samples into the space of the principal components of the Human Genome Diversity Project panel[48,49]. The data was phased using Eagle2 v2.3[50], before imputing with Minimac3 v2.0[51], using a customized reference panel of HRC combined with 2,201 low-coverage whole-genome sequences HUNT samples. Variants with low estimated squared correlations between imputed and true genotypes (R^2 <0.3) were excluded. A logistic regression analysis was run with SAIGE[52], including sex, batch, and 4 PCs as covariates.

23andMe

The 23andMe data consists of 3807 cases and 359,839 controls. Among the controls, there were 19,638 individuals between the age of 45-60 and 340,201 individuals over 60. There were 130 cases between 45-60 and 3677 cases over the age of 60. DNA extraction and genotyping were performed on saliva samples by National Genetics Institute (NGI), a CLIA licensed clinical laboratory and a subsidiary of Laboratory Corporation of America. Samples were genotyped on one of five genotyping platforms. The v1 and v2 platforms were variants of the Illumina HumanHap550+ BeadChip, including about 25,000 custom SNPs selected by 23andMe, with a total of about 560,000 SNPs. The v3 platform was based on the Illumina OmniExpress+ BeadChip, with custom content to improve the overlap with our v2 array, with a total of about 950,000 SNPs. The v4 platform was a fully customized array, including a lower redundancy subset of v2 and v3 SNPs with additional coverage of lower-frequency coding variation, and about 570,000 SNPs. The v5 platform, in current use, is an Illumina Infinium Global Screening Array (~640,000 SNPs) supplemented with ~50,000 SNPs of custom content. Samples that failed to reach 98.5% call rate were re-analyzed.

Only individuals of European ancestry were included in the data. Individuals were assigned ancestry by first partitioning the phased genomic data into short windows of about 300 SNPs. Within each window, a support vector machine (SVM) classified individual haplotypes into one of 31 reference populations (https://www.23andme.com/ancestry-composition-guide/). The SVM classifications are then fed into a hidden Markov model (HMM) that accounts for switch errors and incorrect assignments, and gives probabilities for each reference population in each window. Finally, we used simulated admixed individuals to recalibrate the HMM probabilities so that the reported assignments are consistent with the simulated admixture proportions. Europeans were defined as those with ancestry probabilities of European + Middle Eastern > 0.97 and European > 0.90. Only unrelated individuals were used for the GWAS analysis. Individuals were defined as related if they shared more than 700 cM IBD, including regions where the two individuals share either one or both genomic segments IBD. Cases were preferentially chosen over controls.

Variants were imputed in two separated imputation reference panels. For the first one, we combined the May 2015 release of the 1000 Genomes Phase 3 haplotypes[43] with the UK10K[44] imputation reference panel to create a single unified panel. We used the Human Reference Consortium (HRC) as the second imputation reference panel[42]. Participant data was phased using an internally-developed tool based on Beagle[53] and a new phasing algorithm Eagle[54]. The phased participant data was imputed against both reference panels using Minimac[44]. The resulting imputed data was merged with HRC given preference over the merged panel. The imputed dosage data was analyzed using age, sex, platform, and PCs 1-4 as covariates. The association test *P-value* was computed using a likelihood ratio test.

For QC of genotyped GWAS results, SNPs genotypes only on v1 and v2 platforms were flagged due to low sample size. SNPs on mitochondrial DNA and chromosome Y were flagged. Using trio data, SNPs that failed a test for parent-offspring transmission were flagged (specifically, child's allele count was regressed against the mean parental allele count and flagged SNPs with fitted $\beta<0.6$ and $P<10^{-20}$ for a test of $\beta<1$). SNPs with a Hardy-Weinberg $P<10^{-20}$, or a call rate of <90% were flagged. Genotyped SNPs with batch effects or date effects ($P<10^{-50}$) were flagged. SNPs with large sex effect (ANOVA of SNP genotypes, r2>0.1) were flagged. SNPs with probes matching multiple genomic positions in the reference genome ('self chain') were flagged. For imputed GWAS results, SNPs with Rsq < 0.3, as well as SNPs that had strong evidence of a platform batch effect were flagged. The batch effect test is an F test from an ANOVA of the SNP dosages against a factor representing v4 or v5 platform ($P<10^{-50}$). SNPs with a sample size <20% the total sample were flagged. These flagged SNPs were removed before analysis. Logistic regression results that did not converge due to complete separation, identified by abs(effect)>10 or stderr>10 on the log odds scale were removed. SNPs with MAF < 0.1% were removed.

BioVU

The BioVU data consists of 600 cases and 36,059 controls. Cases were defined as individuals diagnosed with ICD-10 G30 and ICD-9 331.0. Controls were individuals without any of the following ICD-10 diagnoses; G30, F01, F02, F03, F10.27, F10.97, F13.27, F13.97, F18.17, F18.27, F18.97, F19.17, F19.27, F19.97, G31.0, G31.83 and the following ICD-9 diagnoses; 331.0 ,290, 291.2, 292.82, 294.1, 294.10, 294.11, 294.2, 294.20, 294.21, 331.19, 331.82. Individuals with a family history of dementia in their electronic health records were also excluded from the control sample. The participants were genotyped on the Illumina MEGAEX array. The genotypes were filtered for SNP and individual call rates, sex discrepancies, and excessive heterozygosity using PLINK (--geno 0.05, --mind 0.02, |Fhet| > 0.2, HWE $<10\text{x}10^{-10}$). Autosomes were imputed to the HRC panel using Michigan Imputation Server1 in five batches and converted to hardcalls using default PLINK threshold settings.

Non-biallelic SNPs were filtered out and SNPs with imputation quality (R2) less than 0.3. SNPs with minor allele frequency less than 0.005 were removed. SNPs with genotyping rates less than 0.98 were excluded. Individuals with call rates less than 0.98 were excluded.

Principal component analysis (PCA) was used to determine BioVU individuals of European genetic ancestry. First, we performed PCA using FlashPCA[55] on BioVU combined with CEU, YRI, and CHB reference sets from 1000 Genomes Project Phase 3[43]. Principal components were scaled so that the axes could be interpreted as proportions of genetic ancestry. We selected BioVU individuals who were within 40% of the CEU cluster along the CEU-CHB axis and within 30% of the CEU cluster on the CEU-YRI axis, generating a once-PCA filtered European set. To ensure subsequent steps would remove SNPs associated with reduced quality rather than cryptic population substructure, we filtered the previously identified BioVU European cluster to identify individuals falling within the CEU, TSI, and GIH 1000 genomes populations, producing a twice-filtered European set. Using the twice-filtered European set we conducted a series of SNP checks. We filtered individuals with IBS greater than 0.2. We checked for imputation batch effects by conducting pairwise logistic regression of the five imputation batches using sex and top 10 principal components as covariates. SNPs with p-values less than 0.001 in the additive model were flagged. Any SNPs with a MAF difference greater than 0.1 between BioVU and CEU were flagged. SNPs with a Hardy-Weinberg Equilibrium P-value less than $10x10^{-10}$ were flagged. Flagged SNPs were removed from the once-filtered European set. Individuals in the once-filtered European set and SNPs passing QC in the hardcall data were extracted from the dosage data. Finally, a GWAS was performed using SAIGE[52] with default settings including sex and the top 10 PCs as covariates.

DemGene, TwinGene, STSA, Gothenburg, and ANMerge

The origin of the DemGene (1638 cases and 6059 controls), STSA (320 cases and 750 controls), and TwinGene (224 cases and 6321 controls) samples has been previously described in Jansen *et al.* (2019). For the STSA data, informed consent was obtained from all participants and the studies were approved by the Regional Ethics Board in Stockholm and the Institutional Review Board at the University of Southern California. For the TwinGene data, written informed consent was obtained from all participants and the study was approved by the Regional Ethics Board in Stockholm. The ANMerge data (366 cases and 259 controls) consists of 3 batches. Batch 1 is available on synapse.org (synapse ID: syn2795014); the origin and genotyping of this data is described in Birkenbihl *et al.* (2021)[56]. Batch 2 and 3 were both genotyped on Illumina HumanOmniExpress-12 v1.0 and were merged after QC and removal of non-EUR individuals. The merged version of batch 2 and batch 3 were put through the same QC pipeline again and the batch associated variants were removed. Batch associated variants ($P<5x10^{-8}$) were identified through assigning the batch 2 individuals as controls and batch 3 individuals as cases and running Plink logistic regression.

The Gothenburg H70 Birth Cohort Studies and Clinical AD from Sweden (Gothenburg) AD cases originate from Sweden and were either collected in memory clinics (in different parts of Sweden) or as a part of two population-based epidemiological studies in Gothenburg; the Prospective Population Study of Women (PPSW) and the Gothenburg Birth Cohort Studies (H70, H85 and 95+), described in detail previously [57–60]. Controls originate from the Gothenburg Birth Cohort Studies and PPSW. Individuals of non-European descent were excluded as part of the QC of the GWAS-data. AD diagnosis was based on National Institute of Neurological and Communicative Disorders and Stroke-Alzheimer's Disease and Related Disorders (NINCDS-ADRA) criteria. All control samples were clinically investigated and free from dementia. The individuals were genotyped using the Illumina Neurochip array.

Initially, the genotype data were obtained in Plink v1.90[27] binary format and, if necessary, were converted to build GRCh37 using the UCSC LiftOver tool[28]. The raw genotypes were processed using the Psychiatric Genomics Consortium (PGC) Ricopili pipeline version 2019_Aug_16.001. The quality control (QC) procedure initially removed SNPs with a missingness > 0.95, then kept individuals with a SNP missingness < 0.05 and an autosomal heterozygosity deviation (F_{het}) < 0.2. Finally, SNPs with a missingness > 0.02, a difference in SNP missingness between cases and controls > 0.02; and deviation from Hardy-Weinberg equilibrium ($P < 10^{-6}$ in controls or $P < 10^{-10}$ in cases) were removed.

Next, non-European individuals within the datasets were removed based on principle component analysis (PCA), using the 1KG Phase 3 dataset as a reference[29]. The PCA pipeline was repeated including all European individuals in all genotype level datasets to identify individuals across the datasets with a pihat > 0.2 for exclusion from the analysis. PCA was additionally performed within each European dataset to create principal component covariates for logistic regression. The genotype data of the European individuals were imputed to the Haplotype Reference Consortium reference (HRC r1.1 2016)[30] using the Michigan Imputation Server[31]. The imputed genotypes within each dataset were analysed using Plink[27] logistic regression adjusted for covariates. The covariates were principal components 1-4, plus principal components significantly associated with the phenotype, and sex. The significant principal components were identified by testing the first 20 principal components for phenotype association and evaluating their impact on the genome-wide test statistics using λ. At the time of analysis age information was not available for use as a covariate however the cases and controls of each dataset were well matched in age (**Supplementary Table 16**). Unfortunately, only a fraction of age information for DemGene participants (40.4% of cases and 9.19% of controls) was available so age matching cannot be determined.

IGAP

The summary statistics from the International Genomics of Alzheimer's Project (IGAP)[4] were obtained from https://www.niagads.org/datasets/ng00075. The stage 1 results were used in the meta-analysis. The stage 1 results were derived from genotyped and imputed data (11,480,632 variants, phase 1 integrated release 3, March 2012) of 21,982 Alzheimer's disease cases and 41,944 cognitively normal controls. Further information on the methods for generating the summary statistics and phenotyping are available in Kunkle *et al.* (2019)[38]. The data was generated using standard QC procedures. The genotypes were imputed to the 1KG reference panel[43], analyzed with general linear mixed effects models and then meta-analyzed with METAL[37]. Written informed consent was obtained from study participants or, for those with substantial cognitive impairment, from a caregiver, legal guardian or other proxy, and the study protocols for all populations were reviewed and approved by the appropriate institutional review boards.

Finngen

The summary statistics for 1798 cases and 72206 controls from Finngen were obtained from https://storage.googleapis.com/finngen-public-data-r3/summary_stats/finngen_r3_AD_LO_EXMORE.gz. The genotype data were quality controlled with a standard protocol, imputed to SISu v3 reference panel, and analysed using SAIGE[52]. Thorough documentation of data sourcing and processing is available at https://finngen.gitbook.io/documentation/. Cases were defined as being diagnosed with ICD-10 G301, further information regarding the phenotype is available at https://risteys.finngen.fi/phenocode/AD_LO.

GR@CE

The GR@CE data from Moreno-Grau *et al.* (2020)[61] was obtained through the GWAS catalog portal (ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/Moreno-GrauS_31473137_GCST009020/GRACE_StageI.txt). The phenotype was determined through structured neurological evaluation. The data was quality controlled using standard procedures, the genotypes were imputed to the HRC reference panel[42], and the dosages were analysed using an additive model in PLINK v1.9[47] with the top 4 PCs as covariates. Further information is available in Moreno-Grau *et al.* (2020)[61].

Brain regional gene expression

The per-region mean gene expression of the genes that map to the genomic risk loci based on eQTL expression was calculated using GAMBA (alpha version)[31]. A full description of the methods of GAMBA is available in Wei *et al.* (2021)[62]. In short, the gene expression data was obtained from the Allen Human Brain Atlas (http://human.brain-map.org) and the tissue samples were mapped to 64 FreeSurfer cortical and subcortical brain regions to generate a mean regional expression for each gene. Linear regression was used to compare the regional expression of the tested gene-set and the regional expression of the null model gene-set. The tested gene-set included 329 genes that mapped to genomic risk loci based on eQTLs and were present in the GAMBA gene expression data. The brain gene null model gene-set was composed of 329 randomly selected genes which are significantly over-expressed in the brain compared to other GTEx tissues. The random gene model was composed of 329 randomly selected genes with regional gene expression values in GAMBA.

Polygenic risk score

We generated the polygenic risk scores (PRS) using PRSice[63]. The meta-analysis summary statistics used for prediction were generated using a METAL[37] inverse-variance weighted meta-analysis. The Gothenburg dataset was excluded from the METAL meta-analysis to act as an independent test set because it was the dataset with the largest number of cases. The UKB data was excluded from the METAL meta-analysis because the effect estimates of this dataset were calculated for a quantitative phenotype, whereas the effect estimates of all other datasets were calculated for a case-control phenotype. It is not appropriate to include datasets with phenotypes on different scales in an inverse-variance weighted analysis so the UKB data was excluded. The datasets were quality controlled using the same procedures as the larger mvGWAMA meta-analysis (described above). Clumping of the summary statistics was performed using PRSice with clumping of regions <250kb apart and in at least low LD (r^2>0.1). The Gothenburg dataset was quality controlled, imputed, and converted to hard-called genotypes as described above before being utilized as the independent test set. The following P-value thresholds were tested 1, 0.5-0.0001 (by intervals of 0.0001), 5.005×10^{-5}, and 5×10^{-8}. Significantly associated principle components and sex were used as covariates in the PRS models. Versions of the PRS analyses were performed without the larger *APOE* region (GRCh37: 19:40000000-50000000). The PRS utilizing UKB data was performed in the same way as the meta-analysis PRS. To estimate the variance explained by a combination of the UKB and meta-analysis PRS, base model PRS (no covariates) were generated and a regression was performed to estimate the variance explained by the model. The regression included the Gothenburg case-control status as the outcome variable and the individual PRS for each model (UKB and meta-analysis) as predictors.

Enrichment analyses using replicated loci

In order to perform gene-set enrichment analysis with just the significant loci identified in Kunkle *et al.* (2019)[38] and non-significant loci, the following genomic regions (GRCh37) were removed:
AGRN (1:985377-1057677), *NCK2* (2:106122777-106235428), *CLNK* (4:11014822-11044972), *TNIP1* (5:150432388-150432388), *HAVCR2* (5:156506344-156547031), *MHC* (6:32180146-32713511), *TMEM106B* (7:12233848-12285140), *SHARPIN* (8:145018354-145158607), *USP6NL/ECHDC3* (10:11487834-11723537), *CCDC6* (10:61629823-61785671), *ADAM10* (15:58838575-59272096), *APH1B* (15:63441242-63595878), *SCIMP/RABEP1* (17:4958842-5013491), *GRN* (17:42430244-42590812), *ABI3* (17:47297297-47475549), *TSPOAP1-AS1* (17:56398006-56410041), *ACE* (17:61545779-61578207), *NTN5* (19:49168942-49252574), *CD33* (19:51710654-51737991), *LILRB2* (19:54814234-54834217), *APP* (21:27473875-27563105). The MHC region was removed in order to compare to the results from the full analysis. The analyses were performed with the same specifications as the full gene-set analyses, with the exception that the additional analyses with the larger *APOE* region excluded was not performed. The active chromatin and functional consequence enrichment analyses were performed with the same specifications as the full analyses with the previously mentioned genomic regions excluded.

References

1. Jansen, I. E. *et al.* Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat. Genet.* **51**, 404–413 (2019).
2. Ciani, M., Benussi, L., Bonvicini, C. & Ghidoni, R. Genome Wide Association Study and Next Generation Sequencing: A Glimmer of Light Toward New Possible Horizons in Frontotemporal Dementia Research. *Front. Neurosci.* **13**, 506 (2019).
3. Zheng, J. *et al.* LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* **33**, 272–279 (2016).
4. Lambert, J.-C. *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.* **45**, 1452–1458 (2013).
5. Sudlow, C. *et al.* UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLOS Med.* **12**, e1001779 (2015).
6. Kundaje, A. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
7. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164–e164 (2010).
8. MacDonald, R. *et al.* A Novel Egr-1-Agrin Pathway and Potential Implications for Regulation of Synaptic Physiology and Homeostasis at the Neuromuscular Junction. *Front. Aging Neurosci.* **9**, 258 (2017).
9. Rauch, S. M. *et al.* Changes in brain β-amyloid deposition and aquaporin 4 levels in response to altered agrin expression in mice. *J. Neuropathol. Exp. Neurol.* **70**, 1124–1137 (2011).
10. Marques-Coelho, D. *et al.* Differential transcript usage unravels gene expression alterations in Alzheimer's disease human brains. *npj Aging Mech. Dis.* **7**, 2 (2021).
11. Parmar, A. S. *et al.* Association study of FUT2 (rs601338) with celiac disease and inflammatory bowel disease in the Finnish population. *Tissue Antigens* **80**, 488–493 (2012).
12. Verma, A. *et al.* eMERGE Phenome-Wide Association Study (PheWAS) identifies clinical associations and pleiotropy for stop-gain variants. *BMC Med. Genomics* **9 Suppl 1**, 32 (2016).
13. Ramos, C. J. & Antonetti, D. A. The role of small GTPases and EPAC-Rap signaling in the regulation of the blood-brain and blood-retinal barriers. *Tissue barriers* **5**, e1339768 (2017).
14. Sweeney, M. D., Sagare, A. P. & Zlokovic, B. V. Blood–brain barrier breakdown in Alzheimer disease and other neurodegenerative disorders. *Nat. Rev. Neurol.* **14**, 133–150 (2018).
15. Smith, R. G. *et al.* Meta-analysis of epigenome-wide association studies in Alzheimer's disease highlights 220 differentially methylated loci across cortex. *bioRxiv* 2020.02.28.957894 (2020) doi:10.1101/2020.02.28.957894.
16. Yamagishi, S. *et al.* Netrin-5 is highly expressed in neurogenic regions of the adult brain. *Front. Cell. Neurosci.* **9**, 146 (2015).
17. Tan, M.-S., Yu, J.-T. & Tan, L. Bridging integrator 1 (BIN1): form, function, and Alzheimer's disease. *Trends Mol. Med.* **19**, 594–603 (2013).
18. Gratuze, M., Leyns, C. E. G. & Holtzman, D. M. New insights into the role of TREM2 in Alzheimer's disease. *Mol. Neurodegener.* **13**, 66 (2018).
19. Asanomi, Y. *et al.* A rare functional variant of SHARPIN attenuates the inflammatory response and associates with increased risk of late-onset Alzheimer's disease. *Mol. Med.* **25**, 20 (2019).
20. Hawrylycz, M. J. *et al.* An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* **489**, 391–399 (2012).

21. Yin, R.-H., Yu, J.-T. & Tan, L. The Role of SORL1 in Alzheimer's Disease. *Mol. Neurobiol.* **51**, 909–918 (2015).
22. De Roeck, A., Van Broeckhoven, C. & Sleegers, K. The role of ABCA7 in Alzheimer's disease: evidence from genomics, transcriptomics and methylomics. *Acta Neuropathol.* **138**, 201–220 (2019).
23. Schwartzentruber, J. *et al.* Genome-wide meta-analysis, fine-mapping and integrative prioritization implicate new Alzheimer's disease risk genes. *Nat. Genet.* **53**, 392–402 (2021).
24. Dourlen, P., Kilinc, D., Malmanche, N., Chapuis, J. & Lambert, J.-C. The new genetic landscape of Alzheimer's disease: from amyloid cascade to genetically driven synaptic failure hypothesis? *Acta Neuropathol.* **138**, 221–236 (2019).
25. Ghanbari, M. *et al.* A functional variant in the miR-142 promoter modulating its expression and conferring risk of Alzheimer disease. *Hum. Mutat.* **40**, 2131–2145 (2019).
26. Jun, G. R. *et al.* Transethnic genome-wide scan identifies novel Alzheimer's disease loci. *Alzheimer's Dement.* **13**, 727–738 (2017).
27. Liu, S. *et al.* A clinical dose of angiotensin-converting enzyme (ACE) inhibitor and heterozygous ACE deletion exacerbate Alzheimer's disease pathology in mice. *J. Biol. Chem.* **294**, 9760–9770 (2019).
28. Zhao, L. CD33 in Alzheimer's Disease – Biology, Pathogenesis, and Therapeutics: A Mini-Review. *Gerontology* **65**, 323–331 (2019).
29. Karch, C. M. & Goate, A. M. Alzheimer's disease risk genes and mechanisms of disease pathogenesis. *Biol. Psychiatry* **77**, 43–51 (2015).
30. Beck, T. N., Nicolas, E., Kopp, M. C. & Golemis, E. A. Adaptors for disorders of the brain? The cancer signaling proteins NEDD9, CASS4, and PTK2B in Alzheimer's disease. *Oncoscience* **1**, 486–503 (2014).
31. Wei, Y. *et al.* Genetic mapping and evolutionary analysis of human-expanded cognitive networks. *Nat. Commun.* **10**, 4839 (2019).
32. Fama, R. & Sullivan, E. V. Thalamic structures and associated cognitive functions: Relations with age and aging. *Neurosci. Biobehav. Rev.* **54**, 29–37 (2015).
33. Margulies, D. S. *et al.* Mapping the functional connectivity of anterior cingulate cortex. *Neuroimage* **37**, 579–588 (2007).
34. Uddin, L. Q., Nomi, J. S., Hébert-Seropian, B., Ghaziri, J. & Boucher, O. Structure and Function of the Human Insula. *J. Clin. Neurophysiol.* **34**, 300–306 (2017).
35. Smith, K. S., Tindell, A. J., Aldridge, J. W. & Berridge, K. C. Ventral pallidum roles in reward and motivation. *Behav. Brain Res.* **196**, 155–167 (2009).
36. Tate, M. C., Herbet, G., Moritz-Gasser, S., Tate, J. E. & Duffau, H. Probabilistic map of critical functional regions of the human cerebral cortex: Broca's area revisited. *Brain* **137**, 2773–2782 (2014).
37. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
38. Kunkle, B. W. *et al.* Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. *Nat. Genet.* **51**, 414–430 (2019).
39. Gudbjartsson, D. F. *et al.* Large-scale whole-genome sequencing of the Icelandic population. *Nat. Genet.* **47**, 435–444 (2015).
40. Steinthorsdottir, V. *et al.* Identification of low-frequency and rare sequence variants associated with elevated or reduced risk of type 2 diabetes. *Nat. Genet.* **46**, 294–298 (2014).
41. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
42. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
43. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).

44. Huang, J. *et al.* Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel. *Nat. Commun.* **6**, 8111 (2015).
45. Guo, Y. *et al.* Illumina human exome genotyping array clustering and quality control. *Nat. Protoc.* **9**, 2643–2662 (2014).
46. Kent, W. J. BLAT - The BLAST-like alignment tool. *Genome Res.* **12**, 656–664 (2002).
47. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, (2015).
48. Jun, L. Z. *et al.* Worldwide Human Relationships Inferred from Genome-Wide Patterns of Variation. *Science (80-.).* **319**, 1100–1104 (2008).
49. Wang, C. *et al.* Ancestry estimation and control of population stratification for sequence-based association studies. *Nat. Genet.* **46**, 409–415 (2014).
50. Loh, P. R. *et al.* Reference-based phasing using the Haplotype Reference Consortium panel. *Nat. Genet.* **48**, 1443–1448 (2016).
51. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).
52. Zhou, W. *et al.* Scalable generalized linear mixed model for region-based association tests in large biobanks and cohorts. *Nat. Genet.* **52**, 634–639 (2020).
53. Browning, S. R. & Browning, B. L. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* **81**, 1084–1097 (2007).
54. Loh, P.-R., Palamara, P. F. & Price, A. L. Fast and accurate long-range phasing in a UK Biobank cohort. *Nat. Genet.* **48**, 811–816 (2016).
55. Abraham, G., Qiu, Y. & Inouye, M. FlashPCA2: principal component analysis of Biobank-scale genotype datasets. *Bioinformatics* **33**, 2776–2778 (2017).
56. Birkenbihl, C. *et al.* ANMerge: A Comprehensive and Accessible Alzheimer's Disease Patient-Level Dataset. *J. Alzheimers. Dis.* **79**, 423–431 (2021).
57. Skoog, I. *et al.* Decreasing prevalence of dementia in 85-year olds examined 22 years apart: the influence of education and stroke. *Sci. Rep.* **7**, 6136 (2017).
58. Skoog, J. *et al.* A Longitudinal Study of the Mini-Mental State Examination in Late Nonagenarians and Its Relationship with Dementia, Mortality, and Education. *J. Am. Geriatr. Soc.* **65**, 1296–1300 (2017).
59. Kern, S. *et al.* Prevalence of preclinical Alzheimer disease: Comparison of current classification systems. *Neurology* **90**, e1682–e1691 (2018).
60. Zettergren, A. *et al.* The ACE Gene Is Associated with Late-Life Major Depression and Age at Dementia Onset in a Population-Based Cohort. *Am. J. Geriatr. Psychiatry* **25**, 170–177 (2017).
61. Moreno-Grau, S. *et al.* Genome-wide association analysis of dementia and its clinical endophenotypes reveal novel loci associated with Alzheimer's disease and three causality networks: The GR@ACE project. *Alzheimer's Dement.* **15**, 1333–1347 (2019).
62. Wei, Y. *et al.* Statistical testing and annotation of gene transcriptomic-neuroimaging associations. *bioRxiv* 2021.02.22.432228 (2021) doi:10.1101/2021.02.22.432228.
63. Euesden, J., Lewis, C. M. & O'Reilly, P. F. PRSice: Polygenic Risk Score software. *Bioinformatics* **31**, 1466–1468 (2015).

Further acknowledgements

The Trøndelag Health Study (The HUNT Study) is a collaboration between HUNT Research Centre (Faculty of Medicine and Health Sciences, NTNU, Norwegian University of Science and Technology), Trøndelag County Council, Central Norway Regional Health Authority, and the Norwegian Institute of Public Health.
The genetic investigations of the HUNT Study, is a collaboration between researchers from the K.G. Jebsen center for genetic epidemiology and University of Michigan Medical School and the University of Michigan School of Public Health. The K.G. Jebsen Center for Genetic Epidemiology is financed by Stiftelsen Kristian Gerhard Jebsen; Faculty of Medicine and Health Sciences, NTNU, Norway. The genotyping was financed by the National Institute of health (NIH), University of Michigan, The Norwegian Research council, and Central Norway Regional Health Authority and the Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology (NTNU). The genotype quality control and imputation has been conducted by the K.G. Jebsen center for genetic epidemiology, Department of public health and nursing, Faculty of medicine and health sciences, Norwegian University of Science and Technology (NTNU).

AZ was supported by the Swedish Alzheimer Foundation (AF-939988, AF-930582, AF-646061, AF-741361), and the Dementia Foundation (2020-04-13, 2021-04-17).
InSk was supported by the Swedish state under the agreement between the Swedish government and the county councils, the ALF-agreement (ALF 716681), the Swedish Research Council (no 11267, 825-2012-5041, 2013-8717, 2015-02830, 2017-00639, 2019-01096), Swedish Research Council for Health, Working Life and Welfare (no 2001-2646, 2001-2835, 2001-2849, 2003-0234, 2004-0150, 2005-0762, 2006-0020, 2008-1229, 2008-1210, 2012-1138, 2004-0145, 2006-0596, 2008-1111, 2010-0870, 2013-1202, 2013-2300, 2013-2496), Swedish Brain Power, Hjärnfonden, Sweden (FO2016-0214, FO2018-0214, FO2019- 0163), the Alzheimer's Association Zenith Award (ZEN-01-3151), the Alzheimer's Association Stephanie B. Overstreet Scholars (IIRG-00-2159), the Alzheimer's Association (IIRG-03-6168, IIRG-09-131338) and the Bank of Sweden Tercentenary Foundation.
SK was supported by the Swedish state under the agreement between the Swedish government and the county councils, the ALF-agreement (ALFGBG-81392, ALF GBG-771071), the Swedish Alzheimer Foundation (AF-842471, AF-737641, AF-939825), and the Swedish Research Council (2019-02075).
MW was supported by the Swedish Research Council 2016-01590.
HZ is a Wallenberg Scholar supported by grants from the Swedish Research Council (2018-02532), the European Research Council (681712), Swedish State Support for Clinical Research (ALFGBG-720931), the Alzheimer Drug Discovery Foundation (ADDF), USA (201809-2016862), the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 860197 (MIRIADE), and the UK Dementia Research Institute at UCL.
KB was supported by the Swedish Research Council (2017-00915) and the Swedish state under the agreement between the Swedish government and the County Councils, the ALF-agreement (ALFGBG-715986). We would like to thank UCL Genomics, London, UK, for performing the genotyping analyses of the samples within the Gothenburg H70 Birth Cohort Studies and Clinical AD Sweden.
IKK receives funding from Swedish Research Council for Health, Working Life and Welfare (2018-01201) and the National Institutes of Health R01AG060470.

NLP receives funding from the National Institutes of Health Grants No. R01AG059329, R01AG060470. CAR receives funding from the National Institutes of Health (NIH) including RF1AG058068, R01AG060470, R01AG059329, R01AG050595, and R01AG046938.
AH funding acknowledgement as per Kunkle, B.W., et al., Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates A beta, tau, immunity and lipid processing (vol 51, pg 414, 2019). Nature Genetics, 2019. 51(9): p. 1423-1424.
DH was supported by NIH Grant Number U24DA041123
RD is supported by the following: (1) NIHR Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London, London, UK; (2) Health Data Research UK, which is funded by the UK Medical Research Council, Engineering and Physical Sciences Research Council, Economic and Social Research Council, Department of Health and Social Care (England), Chief Scientist Office of the Scottish Government Health and Social Care Directorates, Health and Social Care Research and Development Division (Welsh Government), Public Health Agency (Northern Ireland), British Heart Foundation and Wellcome Trust; (3) The BigData@Heart Consortium, funded by the Innovative Medicines Initiative-2 Joint Undertaking under grant agreement No. 116074. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA; it is chaired by DE Grobbee and SD Anker, partnering with 20 academic and industry partners and ESC; (4) the National Institute for Health Research University College London Hospitals Biomedical Research Centre; (5) the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London; (6) the UK Research and Innovation London Medical Imaging & Artificial Intelligence Centre for Value Based Healthcare; (7) the National Institute for Health Research (NIHR) Applied Research Collaboration South London (NIHR ARC South London) at King's College Hospital NHS Foundation Trust.
JMS is supported by NIH F31MH124306.

This work was funded by The Netherlands Organization for Scientific Research (NWO VICI 453-14-005), NWO Gravitation: BRAINSCAPES: A Roadmap from Neurogenetics to Neurobiology (Grant No. 024.004.012), and a European Research Council advanced grant (Grant No, ERC-2018-AdG GWAS2FUNC 834057). The analyses were carried out on the Genetic Cluster Computer, which is financed by the Netherlands Scientific Organization (NWO: 480-05-003), by the VU University, Amsterdam, The Netherlands, and by the Dutch Brain Foundation, and is hosted by the Dutch National Computing and Networking Services SurfSARA.

This paper represents independent research partly funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. The views expressed are those of the author and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.